



PT H related antigen, or nucleic acid encoding it  
XX  
XX  
PS Example 6B; Fig 6B; 104pp; English.  
XX  
XX This partial protein sequence represents a region of the human  
CC tumour-associated complement factor H (CFH). This sequence is used  
CC in the identification of complement factor H related proteins and  
CC antigens isolated from clone PRB9FH410 (see W39155). The detection of  
CC such proteins and a CFH antigen can be used in screening or for the  
CC treatment of renal or urogenital cancer, e.g. bladder, cervical or  
CC prostate cancer. Agents that may modulate this antigen could be used in  
CC the manufacture of a medicament for the treatment of a tumour cell.  
XX  
SQ Sequence 240 AA.

Query Match 72.6%; Score 1362; DB 18; Length 240;  
Best Local Similarity 100.0%; Pred. No. 9.5e-96;  
Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 78 FFLTGNVFEYGVKAVYTCNEGTLGETINRECDTGDWNTDIPICEVVKCLPVTAPENG 137  
DB 1 flltgnvfevgvkavylcneqyqligelnrecdtdgwtndipicevkcclpvtapeng 60  
QY 138 KIVSSAMEPDREYHFGQAVRFVCONSGYKTEGDEMHCSDDGFWSKPKCVELISCKSPDY 197  
DB 61 kivsamedpreyhfgqavrfvconsgyklegdehmcsddgfwskepkcvelisckspdy 120  
QY 198 INGSPISSQKIIYKENDREFOYKCNMGYEYSEKDAVCTESGMRPLPSCCEKSCDNPYIPNG 257  
DB 121 ingspisqkiiykenderfoycnmgyeysesekdavgctesgmrlpssccekscdnpyipng 180  
QY 258 DVSPIRIKRRTDELTITOCORNGEYPATRGNTAKCTSTGIPAPRCLTKACDIPDIXHGL 317  
DB 181 dyspirikrrtdeeltitocorngeypatrgntaktstgipaprccltkacdipdxhgl 240

RESULT 2  
W39155  
ID W39155 standard; Protein: 216 AA.  
XX  
AC W39155;  
XX  
DT 27-APR-1998 (first entry)  
XX  
DE Clone PRB9FH410 CFH related protein fragment.  
XX  
KW Complement factor H; tumour associated antigen; renal cancer;  
KW urogenital cancer; medicament; modulator.  
XX  
OS Synthetic.  
XX  
PN W09738136-A1.  
XX  
PD 16-OCT-1997.  
XX  
PE 09-APR-1997; 97WO-US05710.  
XX  
PR 06-MAR-1997; 97US-0812481.  
PR 09-APR-1996; 96US-0015083.  
PR 09-APR-1996; 96US-0630048.  
PR 06-MAR-1997; 97US-0038614.  
XX  
PA (BARD-) BARD DIAGNOSTIC SCI INC.  
XX  
PI Enfield DL, Hass GM, Kinders RJ;  
XX  
DR WPI; 1997-512742/47.  
XX  
DR N-PSDB; V02791.  
XX  
PT Treating or screening for cancer, e.g. renal or urogenital cancer -  
PT by modulating or detecting tumour associated human complement factor  
PT H related antigen, or nucleic acid encoding it

XX  
PS Example 6B; Fig 6B; 104pp; English.  
XX  
XX This partial protein is found in clone PRB9FH410 and represents a  
CC complement factor H related protein with homology to a region of the  
CC human tumour-associated complement factor H (CFH). The detection of this  
CC protein and a CFH antigen can be used in screening or for the treatment  
CC of renal or urogenital cancer, e.g. bladder, cervical or prostate cancer.  
CC Agents that may modulate this antigen could be used in the manufacture of  
CC a medicament for the treatment of a tumour cell.  
XX  
SQ Sequence 216 AA.

Query Match 65.5%; Score 1229; DB 18; Length 216;  
Best Local Similarity 99.5%; Pred. No. 9.3e-86;  
Matches 215; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 95 TCNEGTLGETINRECDTGDWNTDIPICEVVKCLPVTAPENGKIVSSAMEPDREYHFGQ 154  
DB 1 tcnegtyllgetlnrecdtdgwtndipicevkcclpvtapengkivsamedpreyhfg 60  
QY 155 AVAFVCONSGYKTEGDEMHCSDDGFWSKPKCVELISCKSPDYINGSPISQKIYKENER 214  
DB 61 avrfvconsgyklegdehmcsddgfwgkpkcvelisckspdyingspisqkiiykener 120  
QY 215 FOYKCNMGYEYSEKDAVCTESGMRPLPSCCEKSCDNPYIPNGDYSPLRIKHRTGDEITY 274  
DB 121 foykcnmgyeysesekdavgctesgmrlpssccekscdnpyipngdysplrikhrtgdeity 180  
QY 275 QCRNGYYPATRGNTAKCTSTGWIAPRCLTKPCDYP 310  
DB 181 qcrngyypatrgntaktstgwiaprccltkpcodyp 216

RESULT 3  
Y55752  
ID Y55752 standard; Protein: 496 AA.  
XX  
AC Y55752;  
XX  
DT 22-FEB-2000 (first entry)  
XX  
DE Human CRI protein LHR-A SCR fragment.  
XX  
KW C3B/C4b receptor; CRI protein; cell-surface protein; erythrocyte; human;  
KW complement regulatory activity; complement pathway enzyme; tissue damage;  
KW reperfusion injury; Arthus reaction; myocardial infarct; inflammation;  
KW heart condition; autoimmune disorder; long homologous repeat; LHR; SCR;  
KW short consensus repeat.  
XX  
OS Homo sapiens.  
XX  
PN US5981481-A.  
XX  
PD 09-NOV-1999.  
XX  
PE 06-JUN-1995; 95US-0470652.  
XX  
PR 03-APR-1989; 89US-0332865.  
PR 06-DEC-1974; 74US-0350238.  
PR 24-FEB-1993; 93US-0026134.  
PR 01-APR-1988; 88US-0176532.  
XX  
PA (UYTO ) UNIV JOHNS HOPKINS.  
PA (BGHM ) BRIGHAM & WOMENS HOSPITAL.  
PA (AVAN-) AVANT IMMUNOTHERAPEUTICS INC.  
XX  
PI Conclino MF, Wong WM, Makrides SC, Klickstein LB, Fearon DT, Ip SH;  
XX  
XX Marsh HC, Carson GR;  
XX  
XX WPI; 1999-633357/54.

Query Match	18.8%	Score 353;	DB 20;	Length 496;
Best Local Similarity	29.8%	Pred. No. 3.7e-19;		
Matches 102;	Conservative 54;	Mismatches 126;	Indels 60;	Gaps 24
3 CNE---	LPRRNRTEILTGSMSDQTYPEGQATYKCRPRGRSLGNVIMVCKRGE-WVALNP	58		
48	capaweliparpnlt-----delefriqlylhyecrpgysgrpfslclcknswtgakd	103		
59	LRRCKRRPCGHGDPDPFGFTTLGNVFEYGYKAVYTCNEGQVLLEINTRC----	DTD 114		
104	--rrrkscrrppdpjrnghvhwkq--lqfsgajkykscetkgylllgs--sataclisgdtv	158		
115	GWTNDIPRICEVVKC-LPVYAPENGKRVSSAMERDRYHNGQAVRFQNSG-----	YKIE 167		
159	lvadnephicdripcqgliprli-tngdfisthre--nlhygsvvlyrcmpsgsggrkwfely	214		
168	GDEEMHC--SD--GHWSEKPKC-WEISKSPDVINGSPIS--QKTIYKENERFQYKCM	221		
215	gepslyctsnddvgylwspapqcllpnkctprpvenqllvsdnrlsfslnevefrcp	274		
222	GVEYSERGAVCTE-SGWRP-LPSCSEKSCDNPYIPN-----GDYSPLRIKHRT	268		
275	gfymkqprlvkcgqalhwkwpelpsc-srycqp--pdvlnhaertgrdkdnfsp-----	324		
269	GDEITTYQRNGRFPATRGMTAKCTSTG-WIP-APRCILPKCD	308		
325	ggevlyscpep-ydlrqaasmrctlpqgdwspaapltecvkscd	365		
RESULT 4				
R13490				
R13490	standard; Protein; 581 AA.			
AC	R13490;			
XX				
DT	30-OCT-1991 (first entry)			
XX				
DE	Human C4 binding protein.			
XX				
XX	C4bp; monomer; complement protein; pJ0D.C4bp.3; SCR;			
KW	short consensus repeat.			
XX				
OS	Homo sapiens.			
XX				
FH	Key			
FT	Peptide			
FT	Protein			
FT	location/Qualifiers			
FT	1..32			
FT	/label= signal_peptide			
FT	33..581			
FT	/label= C4bp			

FT	Region	33..93	/label= SCR8
FT	Region	94..155	/label= SCR7
FT	Region	156..219	/label= SCR6
FT	Region	220..279	/label= SCR5
FT	Region	280..345	/label= SCR4
FT	Region	346..406	/label= SCR3
FT	Region	407..464	/label= SCR2
FT	Region	465..523	/label= SCR1
FT	Domain	524..581	/label= C4bp_core
FT		/note= "responsible for multimer assembly"	
FT	Disulfide-bond	34..80	/note= "intradomain"
FT	Disulfide-bond	65..92	/note= "intradomain"
XX	WO9111461-A.		
XX	08-AUG-1991.		
XX	28-JAN-1991;	91WO-US00567.	
XX	26-JAN-1990;	90US-0470888.	
XX	(BIOG-) BIOGEN INC.		
XX	Pasek MP, Winkler G, Liu TR;		
XX	WPI: 1991-252613/34.		
XX	N-PsDB; Q13242.		
XX	New C4 binding protein fusion proteins and DNA encoding them -		
XX	comprise assemblies of C4bp monomers linked to functional moiety,		
XX	e.g. AZT, useful as delivery vehicles in diagnosis and therapy		
XX			
PS	Example 1; Fig 1: 105pp: English.		
XX	This sequence was deduced from human hepatocyte (Hep G2) cDNA		
XX	obtained following PCR amplification. The protein is a monomer		
XX	containing 8 SCRs. Each SCR forms a looped domain due to the		
XX	presence of two intradomain disulphide bonds (only the disulphide		
XX	bonds of SCR8 are labelled in the Features Table). Within each SCR,		
XX	the first cysteine residue bonds with the third and the second		
XX	cysteine residue bonds with the fourth. This secondary structure is		
XX	responsible for the conformational flexibility of the C4bp monomer.		
XX	The invention covers fusion proteins in which the monomer sequence,		
XX	or a specified subfragment of it, e.g. having 5, 4, 3 or 1 SCR(s)		
XX	is fused to the C-terminal of a protein such as a viral receptor,		
XX	cell ligand, a bacterial, viral or parasitic immunogen, enzyme,		
XX	cytokine, toxin, etc. See also Q13243-51.		
XX	Sequence 581 AA:		
XX			
QY	Query Match	18.8%; Score 353; DB 12; Length 581;	
QY	Best local Similarity	29.3%; Pred. No. 4, 4e-19;	
QY	Matches 100; Conservative	50; Mismatches 141; Indels	50; Gaps
QY	20 SDQTYEPTGTOAIKCRPGY-RSLGNVIMVCRK-GEWVALNPLRKCKRRCGHPGDPPFCT 77		
QY	51 tetrifgtlkytlcypjvyrshstgtlcnsgdewv-yntf--cikykrcripgc----- 102		
QY	78 FTLTGNV-----FEYGVKAVYTCNGGYDLGGINYRCDTD----GWTNDIPICEVVK 128		
QY	103 --lmgvveiktdltsfsgqfiescsgffligsttar-cevqdrfygwnhplpqcivk 159		

```

OY 129 LPVTAPENGKIVSSAMEPDRREYHFGQAVRVCNSGKRIEEDEMHCSD- - - - -GFMSKEK 184
DB 160 kppdltmrhngse- - - - -enfyaygitsvtycdprtsllghaslsctvenetlgvwrpsp 215
OY 185 PKCEVISCSPDINGSPISQ- -KIYKENERFOYKCNMGCEYSEKGDV- -C-TESGWRP 240
DB 216 pceekltcrkpdvshgmsvsgfgrlynykdtlvtfcqkqgfv- - -rgsvlhcdaadskmp 273
OY 241 L-PSCEKSCDN-PYIPNGDYS- -PLRIK- - -HRTGDEITYOCRNCPYPAT- -RGNTAKC- 291
DB 274 sspacecmnscilpdpdphaswelyprtkedyvvgvltlrychpypkykltdeptlvicq 333
OY 292 TSGWIPARCTLKPCDDYDRIKHGILHNMRRP- - - - -YF 327
DB 334 krlwtpryggcealccpepklnngeltqhrksrpanhcvyf 374

RESULT 5
R28570
ID R28570 standard; peptide: 543 AA.
XX
AC R28570;
XX
DE 19-MAR-1993 (first entry)
XX
DE CRI-4 (266-274 KLKQTGNASD) analogue.
XX
KW short consensus repeat; regulator of complement activation;
KW C3b binding; C4b binding; human complement type 1 receptor.
XX
OS Homo sapiens.
XX
FH Key location/Qualifiers
FT Region 1..60
FT /label= SCR-1
FT Region 61..122
FT /label= SCR-2
FT Region 451..510
FT /label= SCR-8
FT Region 511..543
FT /label= SCR-9
FT /note= "TRUNCATED"
FT MISC-difference 266..274
FT /note= "ERTORDKN substituted with KLKQTGNASD
FT from SCR-8-9 to increase C3b binding"
XX
PN EP512733-A.
XX
PD 11-NOV-1992.
XX
XX 28-APR-1992; 92EP-0303826.
XX
XX 03-MAY-1991; 91US-0695514.
XX
PA (UNIM ) UNIV WASHINGTON.
XX
PI Atkinson JP, Hourcade D, Krych M;
XX
DR WPI; 1992-375009/46.
XX
XX
XX Complement activity regulator protein analogues - useful for
XX treating auto-immune diseases, to suppress transplant rejection,
XX for diagnosis etc.
XX
XX Example 8; Page 18 and R11810; 23pp. English.
XX
XX The CDNA clone designated CRI-4 was described in J.Exp.Med.(1988)
XX 168:1255-1270. It encodes the first 8 and a half amino terminal
XX SCRs of CRI. The invention concerns analogues of "regulator of
XX complement activation" proteins or truncated, hybrid or recombinant
XX forms of them. CRI-4 is a preferred truncated form and a number of
XX specified substitution variants of it are disclosed in which certain

```

```

CC positions in SCR-5-6 are substituted by amino acids from
CC the corresponding positions in SCRs which are involved in C3b- and
CC C4b-binding. The substitution variant given here has increased C3b-
CC binding. The specification does not contain the CRI-4 sequence;
CC the sequence given here was constructed from the full-length CRI
CC amino acid sequence having GENESEQ accession number R11810 and
CC descriptions in the disclosure.
XX
SQ Sequence 543 AA:
XX
Query Match 18.8%; Score 352.5; DB 13; Length 543;
Best Local Similarity 29.6%; Pred. No. 4,5e-19;
Matches 101; Conservative 53; Mismatches 128; Indels 59; Gaps 23;
OY 3 CNE- - -LPRRAMEILTGMSDQTYEGGQATIKCKRPGYRSLGNVIMVCRKGWALNPL 59
DB 2 cnapewlpfartnlt- - - - -delfrpdglylnyccrpyggrpfslclksvwtgkd- 56
OY 60 RKGQRKPCGHPGDPTEFTLTGNAVFEYGVAVYTCNNGYQHLGEINREC- - -DTDG 115
DB 57 -rcrrtkscrnpdpvngmwhvlyk- - -lqfsgqkyctctgyrligs-satclisgdtvl 112
OY 116 WTNDIPICEVWC- -LPVTAPENGKIVSSAMEPDRREYHFGQAVRVCNSG- - - - -YKIEG 168
DB 113 wdnefrpdcrlpogllppl- - -tngdflstnre- - - - -fhysvvtlyrcnpgsgqkrfelyg 168
OY 169 DEEMHC- - -SD- - -GFMSKRPKC- -VEISCKSPDVINGSPIS- -QKIYKENERFOYKCNNG 222
DB 169 epslyctsnddqvglysgpapecllpnktcpvnevgllvsdnrslfsinevefrqgv 228
OY 223 YEYSEKDAVCTE- -SGWRP- -LPSCERKSCDNPIY- - - - -NGDYSPLRIKHRTG 269
DB 229 fymkgprrvkcqalhwepelpsc- - -srvcqrp- - -pdvlhakikltqnsdpsp- - - - -g 278
OY 270 DEITYOCRNCFYPATRGNTAKCTSTG- -WIP- -APRCTIKPCD 308
DB 279 qevfyscepg-ydlrtaasmrctpgdwsapaaptcevxscd 318

RESULT 6
W73147
ID W73147 standard; protein: 778 AA.
XX
AC W73147;
XX
DE 29-JAN-1999 (first entry)
XX
XX
XX Amino acid sequence of the soluble complement receptor 1 (SCR1).
XX
XX Human: soluble complement receptor 1; SCR1; T-cell; B-cell;
XX mediated immune response; inhibition; tissue rejection; gene therapy;
XX dystrophin; inflammatory response; interferon-gamma secretory response;
XX auto-immune response; neurological response; Alzheimer's disease;
XX Parkinson's disease; multiple sclerosis; systemic lupus erythematosus;
XX rheumatoid arthritis; myasthenia gravis; epidermis bullousa;
XX Hashimoto's disease.
XX
XX Homo sapiens.
XX
XX W09845430-A1.
XX
XX 15-OCN-1998.
XX
XX 06-APR-1998; 98WO-GB01012.
XX
XX 05-APR-1997; 97GB-0006950.
XX
XX (ANNE/) ANNENKOV A.
XX (CHER/) CHERNAJOVSKY Y.
XX
XX Annenkov A, Chernaiovsky Y;
XX
XX

```

DR WPI: 1998-568350/48.  
 XX Fragment of soluble human complement receptor 1 - useful for  
 PT treating T-cell or B-cell mediated immune responses e.g.  
 PT inflammatory responses such as rheumatoid arthritis  
 PS  
 XX Disclosure; Fig 1; 54pp; English.  
 XX This is an amino acid sequence of the human soluble complement  
 CC receptor 1 (SCR1), useful in the treatment of T-cell or B-cell  
 CC mediated immune responses. It is used to inhibit a T-cell or  
 CC B-cell mediated immune response to prevent immune response-mediated  
 CC tissue rejection and destruction or clearance or inactivation of an  
 CC expressed protein especially from cells that have been treated by gene  
 CC therapy to express the protein, e.g. dystrophin. The protein can also  
 CC be used to inhibit a T-cell or B-cell mediated inflammatory response,  
 CC an interferon-gamma secretory response, autoimmune response or  
 CC neurological response, e.g. Alzheimer's or Parkinson's disease or  
 CC multiple sclerosis. Also the protein can be used to treat systemic  
 CC lupus erythematosus, rheumatoid arthritis, myasthenia gravis,  
 CC epidermis bullous or Hashimoto's disease.  
 CC  
 XX Sequence 778 AA:  
 SO  
 Query Match 18.7%; Score 350.5; DB 19; Length 778;  
 Best Local Similarity 29.3%; Pred. No. 9.7e-19;  
 Matches 100; Conservative 55; Mismatches 127; Indels 59; Gaps 23;  
 QY 3 CNE---LPPRRNTEILTGSMSDQTYPEGTQAIYKCRPGYSLGNVIMVCKRKGWALNPL 59  
 DB 48 cnapewlpfarphnlt-----defefpigtlylneyecrpgysgrpfslclksnsvtgakd- 102  
 QY 60 RKCKRPGCHPGDTPFGTFTLTGNNVEYGVKAVYTCNCEYOLLEINRYREC---DTDG 115  
 DB 103 -rcrkrcrnpdpvngmwhvlykg--lqfsgqkysctkylrlgs-saatcllsgdtvla 158  
 QY 116 WTNDIPICEVVKC-LPVTAPENKIVSAMPEPDREYHFGQAVRFVNSG-----YKIEG 168  
 DB 159 wdnetpictdripclppl-tngdfistnre---nfhygsvlyrcnpgsgqrkfelvg 214  
 QY 169 DEEMHC--SDD--GEWSKEKPKC-VEISCKSPDVINGSPIS-OKIYKENERFOYKCNMG 222  
 DB 215 epslyctsndqygiwspapqclipmkcprpvenngilvsdnrlsinevefricpgp 274  
 QY 223 YEYSERGDVCTE-SGMRP-LPSCSEKSCDNPIYIPN-----GDYSPRIKIHRTG 269  
 DB 275 fvmkgprtrvkqalnwepelpsc-srvccpp--pdvlnhaertgrdkdnfisp-----g 324  
 QY 270 DEITYOCRNNGFYPATRGNTAKCTSTG-WIP-APRCTLKPDC 308  
 DB 325 gevfysecpq-ydlrngaasmrcrpgdwpapactcevkscd 364  
 Db  
 RESULT 7  
 W45899  
 ID W45899 standard; peptide; 1930 AA.  
 XX W45899;  
 XX  
 AC  
 XX 30-JUN-1998 (first entry)  
 DT  
 XX Human complement receptor 1 (residues 1-1929).  
 DE  
 XX Membrane binding element; thrombotic disease; soluble protein;  
 KW complement-related disease; integral membrane protein; inflammation;  
 KW short consensus repeat; SCR 1-3; CRI; complement receptor type 1.  
 XX  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FH 1930  
 FT Cross-Links  
 FT /note= "Disulphide linked to Cys in peptide given

FT in W45889"  
 XX W09802454-A2.  
 XX 22-JAN-1998.  
 PD  
 XX 08-JUL-1997; 97WO-EP03715.  
 PF  
 XX 15-JUL-1996; 96GB-0014871.  
 PR  
 XX (ADPR-) ADPROTECH PLC.  
 PA  
 PI Dodd I, Mossakowska DEL, Smith RAG;  
 DR WPI: 1998-110524/10.  
 XX  
 XX Derivatives of soluble polypeptide(s) bonded to low affinity  
 PT membrane binding groups - useful for treating complement-related and  
 PT thrombotic diseases, providing improved localisation at cellular  
 PT membranes  
 PS  
 XX Claim 22; Pages 60-61; 75pp; English.  
 CC  
 CC This sequence represents human complement receptor 1 (CRI, CD 35)  
 CC N-terminal fragment. The invention relates to a soluble derivative (A)  
 CC of a soluble polypeptide (I), which comprises at least 2 heterologous  
 CC membrane-binding elements (MBE) of low membrane affinity covalently  
 CC associated with (I). MBE interact, independently and with thermodynamic  
 CC additivity, with components of cellular or artificial membranes exposed  
 CC to extracellular fluids. (A) are used to treat disorders treatable with  
 CC (I) itself, specifically inflammation or any other complement-related  
 CC disorder (e.g. neurological disease, graft rejection, myocardial  
 CC infection, sepsis, rheumatoid arthritis and many others; including  
 CC application to indwelling devices) and thrombolytic disease, but also to  
 CC treat allergy, induce weight loss, to treat ischaemia or asthma and as  
 CC immuno-modulators for treating multiple sclerosis. (A) are administered  
 CC orally, topically, by injection or inhalation at 0.01-10 (preferably  
 CC 0.1-10) mg/kg/day.  
 CC  
 XX Sequence 1930 AA:  
 SQ  
 Query Match 18.7%; Score 350.5; DB 19; Length 1930;  
 Best Local Similarity 29.3%; Pred. No. 2.9e-18;  
 Matches 100; Conservative 55; Mismatches 127; Indels 59; Gaps 23;  
 QY 3 CNE---LPPRRNTEILTGSMSDQTYPEGTQAIYKCRPGYSLGNVIMVCKRKGWALNPL 59  
 DB 2 cnapewlpfarphnlt-----defefpigtlylneyecrpgysgrpfslclksnsvtgakd- 56  
 QY 60 RKCKRPGCHPGDTPFGTFTLTGNNVEYGVKAVYTCNCEYOLLEINRYREC---DTDG 115  
 DB 57 -rcrkrcrnpdpvngmwhvlykg--lqfsgqkysctkylrlgs-saatcllsgdtvla 112  
 QY 116 WTNDIPICEVVKC-LPVTAPENKIVSAMPEPDREYHFGQAVRFVNSG-----YKIEG 168  
 DB 113 wdnetpictdripclppl-tngdfistnre---nfhygsvlyrcnpgsgqrkfelvg 168  
 QY 169 DEEMHC--SDD--GEWSKEKPKC-VEISCKSPDVINGSPIS-OKIYKENERFOYKCNMG 222  
 DB 169 epslyctsndqygiwspapqclipmkcprpvenngilvsdnrlsinevefricpgp 228  
 QY 223 YEYSERGDVCTE-SGMRP-LPSCSEKSCDNPIYIPN-----GDYSPRIKIHRTG 269  
 DB 229 fvmkgprtrvkqalnwepelpsc-srvccpp--pdvlnhaertgrdkdnfisp-----g 278  
 QY 270 DEITYOCRNNGFYPATRGNTAKCTSTG-WIP-APRCTLKPDC 308  
 DB 279 gevfysecpq-ydlrngaasmrcrpgdwpapactcevkscd 318  
 Db  
 RESULT 8  
 Y55751

ID	Accession	Protein	Score	DB	Length	Query
XX	Y55751	standard; Protein; 2039 AA.	18.7%	Score 350.5;	DB 20;	Length 2039;
XX	AC	Y55751;	Best Local Similarity	29.3%;	Pred. No. 3.1e-18;	
XX	DT	22-FEB-2000 (first entry)	Matches 100;	Conservative 55;	Mismatches 127;	Indels 59; Gaps 23
XX	DE	Human C3b/C4b receptor (CRL) protein.				
XX	KW	C3B/C4B receptor; CRI protein; cell-surface protein; erythrocyte; human;				
XX	KW	complement regulatory activity; complement pathway enzyme; tissue damage;				
XX	KW	heart condition; autoimmune disorder; diagnostic.				
XX	OS	Homo sapiens.				
XX	PN	US5981481-A.				
XX	PD	09-NOV-1999.				
XX	PF	06-JUN-1995;	9505-0470652.			
XX	PR	03-APR-1989;	89US-0332865.			
XX	PR	26-DEC-1974;	74US-0350238.			
XX	PR	24-FEB-1993;	93US-0026134.			
XX	PR	01-APR-1988;	88US-0176532.			
XX	PA	(UYUO ) UNITV JOHNS HOPKINS.				
XX	PA	(BGHM ) BRIGHAM & WOMENS HOSPITAL.				
XX	PA	(AVAN-) AVANT IMMUNOTHERAPEUTICS INC.				
XX	PI	Concino MF, Wong WM, Makrides SC, Klickstein LB, Pearson DT, Ip SH;				
XX	PI	Marsh HC, Carson GR;				
XX	DR	WPI; 1999-633357/54.				
XX	DR	N-PSDB; 238150.				
XX	PT	A human C3b/C4b receptor (CRL) protein having antiinflammatory and				
XX	PT	cardiant activity -				
XX	PS	Disclosure; Fig 1A-P; 87pp; English.				
XX	CC	The invention relates to a human C3B/C4B receptor (CRI) protein. The CRI				
XX	CC	protein or fragment is expressed as a cell-surface protein on the surface				
XX	CC	of a non human cell and exhibits a complement regulatory activity of full				
XX	CC	-length human CRI as expressed on erythrocytes. The CRI function in vivo				
XX	CC	may be mediated through the inhibition of complement pathway enzymes. The				
XX	CC	soluble CRI protein exhibits a complement regulatory activity, and this				
XX	CC	may be used to prevent reperfusion injury, inhibit Arthus reaction, and				
XX	CC	neutrophil mediated tissue damage, and reduce myocardial infarct size,				
XX	CC	and inflammation. The CRI protein and its fragments can also be used in				
XX	CC	the treatment of conditions which involve unwanted complement activity,				
XX	CC	e.g. shock lung, tissue damage due to burn, or ischemic heart conditions,				
XX	CC	and autoimmune disorders. CRI proteins, analogues, derivatives, and anti				
XX	CC	-CRI antibodies are used in assays, and diagnostics. The present sequence				
XX	CC	represents the human CRI protein.				
XX	Sequence	2039 AA;				
QY	Query Match	18.7%; Score 350.5; DB 20; Length 2039;				
Db	Best Local Similarity	29.3%; Pred. No. 3.1e-18;				
QY	Matches 100; Conservative 55; Mismatches 127; Indels 59; Gaps 23					
QY	3 CNE---LPPRRRTTELTSSMSDQTYPEGRQATLYKRCPRGKRSICNVIMCRKGEWALNPL 59					
Db	43 cnapewlfparpnlt-----defepfiplyllyncrpygsgrpfisclcknsvwigaka- 97					
QY	60 RKQGRPCGHPGDPFEGFTTLTGWVFEGYKAVYTCNEGYYLLGEINREC-----DTDG 115					
Db	98 -rctkscrnppdpvngmwhvlyg--ldfgsqikysckkgyllls-satctliagdtvi 153					
QY	116 WTNDIPICEVAKC-LPVTAPENGKATVSSAMEDPREYHFGQAVFVCSNG-----YKTEG 168					

ID	Accession	Protein Name	Length (AA)	Score	DB	Length
Db	154	wdnetplctdrlpogllppli-tnngdifsture---nfllygsvvlrcpbgsgyrlkwfelvg	209			
Qy	169	DEEMHC--SDD-GFWSKERKRC-VEISCKSPDYINGSPIS-QKLIYKENERFOYCKNMG	222			
Db	210	epslyctsdndqvgiwsqpaqpcllpkktclpnpvenglldvnsdlsflfnevetrlcpqg	269			
Qy	223	YKESERDVACTE-SGWRP-LPSCSEKSCDNPIPN-----GDYSPLRIRKARTG	269			
Db	270	fvmkyprirvkqgalnkxwepelpsc-srvccpp--pdvllaertlqrkdndfisp-----g	319			
Qy	270	DEITYQCRNGEFPATRGNTAKTSTG-WIP-APRCTLRKPCD	308			
Db	320	gevfyscepg-ydlrgaasmrctcpqgdwspapltcevkscd	359			
RESULT	9					
P92219		P92219 standard; protein; 2317 AA.				
XX						
AC		P92219;				
XX						
DT	22-FEB-1990	(first entry)				
XX						
DE		CRI protein.				
XX						
KW		Complement; cofactor.				
XX						
OS		Homo sapiens (human).				
XX						
FT		Key Location/Qualifiers				
FT		Peptide 10..50				
XX		/label= signal_peptide				
PN		W08909220-A.				
XX						
PD		05-OCT-1989.				
XX						
PF		31-MAR-1989; 89WO-US01358.				
XX						
PR		01-APR-1988; 88US-0176532.				
XX						
PA		(TCEL ) T CELL SCIENCES INC.				
PA		(UYUO ) THE JOHNS HOPKINS UNIVERSITY.				
PA		(BRIG ) THE BRIGHAM AND WOMEN'S HOSPITAL.				
XX						
PI		Featon DT, Klickstein LB, Wong W, Carson G, Concino MF, Makrides SC;				
XX						
DR		WPI: 1989-309498/42.				
DR		N-PSDB; N91477.				
XX						
PT		New nucleic acid sequences encoding new CRI protein - and its fragment,				
PT		for diagnosis and control of complement-related immune defects,				
PT		inflammation, myocardial infarct, etc				
XX						
PS		Claim 1; Fig 1; 191pp; English.				
XX						
CC		This is full-length CRI protein, and shortened forms are new, lacking				
CC		the transmembrane region. The proteins and fragments bind C3b and/or				
CC		C4b, have cofactor activity and inhibit C3 and C5 convertase activity.				
CC		In the sequence, x=untranslated region. This has 7 short consensus				
CC		repeats (CR) of 4 long direct homologous repeats (LHR) (see Fig. 10 in				
CC		patent). Deletion mutants, eg LHR-A for binding of C4b, and LHR-C				
CC		for C3b binding, be made, and LHR-B and -C can detect CR2 sequences.				
CC		They are useful in diagnosing and treating immune disorders, and prevent				
XX		perfusion injury.				
XX						
SO		Sequence 2317 AA;				



FT Misc-difference 53 /note= "Thr substituted by Ser (SCR-8)"  
 FT AC R28560; /note= "Gly substituted by Ser (SCR-8)"  
 FT FT Misc-difference 54 /note= "Ala substituted by Pro (SCR-8)"  
 FT EP512733-A.  
 PD 11-NOV-1992.  
 XX 28-APR-1992; 92EP-0303826.  
 XX 03-MAY-1991; 91US-0695514.  
 XX (UNIM ) UNIV WASHINGTON.  
 PA Atkinson JP, Hourcade D, Krych M;  
 DR WPI; 1992-375009/46.  
 XX  
 PT Complement activity regulator protein analogues - useful for  
 PT treating auto-immune diseases, to suppress transplant rejection,  
 PT for diagnosis etc.  
 PS  
 XX  
 PS Claim 11: Fig 2 and R11810; 23pp; English.  
 CC The cDNA clone designated CRI-4 was described in J.Exp.Med.(1988)  
 CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
 CC SCRs of CRI. The invention concerns analogues of "regulator of  
 CC Complement activation" proteins or truncated, hybrid or recombinant  
 CC forms of them. CRI-4 is a preferred truncated form and a number of  
 CC specified substitution variants of it are claimed in which certain  
 CC positions in SCR-1 which have been identified as important for the  
 CC degree of C3b- and C4b-binding are substituted by amino acids from  
 CC the corresponding positions in SCR-8. The specification does not  
 CC contain the CRI-4 sequence; the sequence given here was constructed  
 CC from the full-length CRI amino acid sequence having GENESEQ  
 CC accession number R11810 and descriptions in the disclosure.  
 CC  
 XX  
 XX Sequence 543 AA;

Query Match 18.4%; Score 345.5; DB 13; Length 543;  
 Best Local Similarity 29.3%; Pred. No. 1.5e-18;  
 Matches 100; Conservative 56; Mismatches 126; Indels 59; Gaps 23;  
 QY 3 CNE---LPPRRNTEILGSMDSOTYREGTOAIYKRCRGYSGLGNVWCKRGGEVVALNPL 59  
 DB ||||| : : : : : ||||| : : : : :  
 DB 2 cnapewlfpfarpnl-----defefipgitylneyecrpgysgrfslclcknsw--sspk 55  
 QY 60 RKCQRPCGHPGDPFPGFTFLTGGNVEYGVKAVYTCNEGVOILGELINRYRC-----DTDG 115  
 DB ||||| : : : : : ||||| : : : : :  
 DB 56 drcrkscrrppdvngmwnvikg--lqfsgqlyskctkyrlligs--ssatclisgdtvi 112  
 QY 116 WTNDIPICEVKK-LPYTAPENCKIVSAMPEDEYHFGQAVRFVNSG-----YKIEG 168  
 DB ||||| : : : : : ||||| : : : : :  
 DB 113 wdneplicdrlpcglppli--lmgfistlure---nfhygsvlyrcmpgsggrkvfelvg 168  
 QY 169 DEEMHC--SDD--GFWGKEKPKC-VEIISCKSPDYINSGPIS-QKTIYKKNFRFOYKCNMG 222  
 DB ||||| : : : : : ||||| : : : : :  
 DB 169 epsiylcsndddqyglwspapqcllpnkctcpnvengllvsdnslfsinevefrcp 228  
 QY 223 YEYSERGDAACTE-SGRP-LPSCEEKSCDNPIYIPN-----GDSPLRIKHRTG 269  
 DB ||||| : : : : : ||||| : : : : :  
 DB 229 fvmkgprtrvvcqalnkwpelpsc--stvcqpp--pdvlnhertqrdkdnisp-----g 278  
 QY 270 DETTYOCRNCFYPATRGNTAKCTSTG-WIP-APRCTLKPDC 308  
 DB ||||| : : : : : ||||| : : : : :  
 DB 279 qevfyscepg-ydlrgaasmrctcpqgwpspaaptcevkscd 318

RESULT 12  
 R28560

ID R28560 standard; peptide; 543 AA.  
 XX AC R28560;  
 XX 19-MAR-1993 (first entry)  
 DE CRI-4 (114S) analogue.  
 XX  
 XX Short consensus repeat; regulator of complement activation;  
 KW C3b binding; C4b binding; human complement type 1 receptor.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 1..60  
 FT /label= SCR-1  
 FT Region 61..122  
 FT /label= SCR-2  
 FT Region 451..510  
 FT /label= SCR-8  
 FT Region 511..543  
 FT /label= SCR-9  
 FT /note= "TRUNCATED"  
 FT Misc-difference 114 /note= "Asp substituted by Ser (SCR-9)"  
 FT  
 XX EP512733-A.  
 XX  
 XX 11-NOV-1992.  
 XX  
 PD 28-APR-1992; 92EP-0303826.  
 XX 03-MAY-1991; 91US-0695514.  
 XX (UNIM ) UNIV WASHINGTON.  
 PA Atkinson JP, Hourcade D, Krych M;  
 DR WPI; 1992-375009/46.  
 XX  
 PT Complement activity regulator protein analogues - useful for  
 PT treating auto-immune diseases, to suppress transplant rejection,  
 PT for diagnosis etc.  
 PS  
 XX  
 PS Claim 11: Fig 2 and R11810; 23pp; English.  
 CC The cDNA clone designated CRI-4 was described in J.Exp.Med.(1988)  
 CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
 CC SCRs of CRI. The invention concerns analogues of "regulator of  
 CC Complement activation" proteins or truncated, hybrid or recombinant  
 CC forms of them. CRI-4 is a preferred truncated form and a number of  
 CC specified substitution variants of it are claimed in which certain  
 CC positions in SCR-2 which have been identified as important for the  
 CC degree of C3b- and C4b-binding are substituted by amino acids from  
 CC the corresponding positions in SCR-9. The specification does not  
 CC contain the CRI-4 sequence; the sequence given here was constructed  
 CC from the full-length CRI amino acid sequence having GENESEQ  
 CC accession number R11810 and descriptions in the disclosure.  
 CC  
 XX  
 XX Sequence 543 AA;

Query Match 18.3%; Score 343.5; DB 13; Length 543;  
 Best Local Similarity 29.0%; Pred. No. 2.1e-18;  
 Matches 99; Conservative 56; Mismatches 127; Indels 59; Gaps 23;  
 QY 3 CNE---LPPRRNTEILGSMDSOTYREGTOAIYKRCRGYSGLGNVWCKRGGEVVALNPL 59  
 DB ||||| : : : : : ||||| : : : : :  
 DB 2 cnapewlfpfarpnl-----defefipgitylneyecrpgysgrfslclcknswtga 56  
 QY 60 RKCQRPCGHPGDPFPGFTFLTGGNVEYGVKAVYTCNEGVOILGELINRYRC-----DTDG 115  
 DB ||||| : : : : : ||||| : : : : :  
 DB 57 -rcrikscrrppdvngmwnvikg--lqfsgqlyskctkyrlligs--ssatclisgdtvi 112





Db 229 fvmkgprvkcqalnkwepe1psc-srvccpp--pdv1haertgrdkdnfsp-----g 278  
Oy 270 DEITYOCNRNGFYPATRGNTAKCTSTG-WIP-APRCTLKPCD 308  
| : | | | | : | | | | | : | |  
Db 279 gevfysecpq-ydlrgaasmrctpgqdwspapicevkscd 318

Search completed: November 21, 2000, 16:43:48  
Job time: 287 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2000, 16:49:19 ; Search time 27.94 Seconds  
(without alignments)  
303.014 Million cell updates/sec

Title: US-09-316-163-10

Perfect score: 1497  
Sequence: 1 EDCNELPPRRNTETITGSMS.....EKSCDNPYIPNGDYSPLRIK 265

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 87993 seqs, 31947931 residues

Total number of hits satisfying chosen parameters: 87993

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1497	100.0	1231	1 CFAH_HUMAN	P08603 homo sapien
2	1024	68.4	1234	1 CFAH_MOUSE	P06909 homo musculu
3	303	20.2	597	1 C4BP_HUMAN	P04003 homo sapien
4	301.5	20.1	263	1 VCP_VACCV	P10998 vaccinia vi
5	298.5	19.9	2039	1 CRI_HUMAN	P17927 homo sapien
6	276.5	18.5	558	1 C4BP_MOUSE	P03514 rattus norv
7	275	18.4	459	1 C4BP_MOUSE	P08607 mus musculu
8	274	18.3	377	1 MCP_HUMAN	P15529 homo sapien
9	272	18.2	610	1 LEM2_HUMAN	P16581 homo sapien
10	266	17.8	484	1 LEM2_PIG	P08110 sus scrofa
11	265.5	17.7	610	1 C4BP_BOVIN	Q28065 bos taurus
12	264.5	17.7	360	1 CCBP_HSVSA	Q01016 herpesvirus
13	262	17.5	381	1 DAF_HUMAN	P08174 homo sapien
14	260.5	17.4	768	1 LEM3_MOUSE	O01102 mus musculu
15	260	17.4	612	1 LEM2_MOUSE	O00690 mus musculu
16	257.5	17.2	1025	1 CR2_MOUSE	P19070 mus musculu
17	254	17.0	768	1 LEM3_RAT	P08106 rattus norv
18	254	17.0	1033	1 CR2_HUMAN	P20023 homo sapien
19	253	16.9	345	1 APOH_MOUSE	Q01333 mus musculu
20	252	16.8	340	1 DAF_PONPY	P49457 pongo pygma
21	252	16.8	661	1 F13B_HUMAN	P05160 homo sapien
22	250.5	16.7	830	1 LEM3_HUMAN	P16109 homo sapien
23	249.5	16.7	769	1 LEM3_SHEEP	P08109 ovis aries
24	246.5	16.5	507	1 DAF_CAVPO	Q06401 cavia porce
25	241	16.1	549	1 LEM2_RAT	P08105 rattus norv
26	241	16.1	611	1 LEM2_CANFA	P33730 canis fami
27	240	16.0	345	1 APOH_BOVIN	P17690 bos taurus
28	239.5	16.0	551	1 LEM2_RABIT	P27113 oryctolagus
29	234	15.6	668	1 F13B_MOUSE	O09168 mus musculu
30	233.5	15.6	958	1 HIG_DROME	O09101 drosophila
31	231.5	15.5	390	1 APOH_MOUSE	Q01475 mus musculu
32	231	15.4	345	1 APOH_HUMAN	P02749 homo sapien
33	222.5	14.9	407	1 DAF2_MOUSE	Q01476 mus musculu

34	222.5	14.9	485	1 LEM2_BOVIN	P08107 bos taurus
35	222	14.8	345	1 APOH_CANFA	P33703 canis fami
36	216.5	14.5	646	1 LEM3_BOVIN	P42201 bos taurus
37	210	14.0	372	1 LEM1_MOUSE	P18337 mus musculu
38	203.5	13.6	372	1 LEM1_HUMAN	P14151 homo sapien
39	203.5	13.6	372	1 LEM1_PANTR	O95237 pan troglod
40	202.5	13.5	372	1 LEM1_PONPY	O95235 pongo pygma
41	200	13.4	372	1 LEM1_RAT	P30836 rattus norv
42	194	13.0	372	1 LEM1_MACMU	Q95198 macaca mula
43	194	13.0	372	1 LEM1_PAPHA	Q28768 papio hamad
44	192.5	12.9	270	1 CFEH_HUMAN	P36980 homo sapien
45	192	12.8	297	1 APOH_RAT	P26644 rattus norv

## ALIGNMENTS

RESULT 1  
ID CFAH\_HUMAN STRAND: PRT: 1231 AA.  
AC P08603;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE COMPLEMENT FACTOR H PRECURSOR.  
GN HFI OR HF OR CFH.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
RN [1]  
RP TISSUE=LIVER;  
RC SEQUENCE FROM N.A.  
RX MEDLINE: 86134059.  
RA Ripoché J., Day A.J., Harris T.J.R., Sim R.B.;  
RT "The complete amino acid sequence of human complement factor H.";  
RL Biochem. J. 249:593-602(1988).  
RN [2]  
RP SEQUENCE OF 53-445 FROM N.A.  
RX MEDLINE: 87054207.  
RA Schultz T.F., Schwaebler W., Stanley K.K., Weiss E., Dierich M.P.;  
RT "Human complement factor H: Isolation of cDNA clones and partial cDNA  
sequence of the 38-kDa tryptic fragment containing the binding site  
for C3b.";  
RL Eur. J. Immunol. 16:1351-1355(1986).  
RN [3]  
RP SEQUENCE OF 226-445 FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE: 86169701.  
RA Kristensen T., Wetsel R.A., Tack B.F.;  
RT "Structural analysis of human complement protein H: homology with C4b  
binding protein, beta 2-glycoprotein I, and the Ba fragment of B2.";  
RL J. Immunol. 136:3407-3411(1986).  
RN [4]  
RP SEQUENCE OF 1047-1231 FROM N.A.  
RX MEDLINE: 91201892.  
RA Estallier C., Kolsteden V., Schwaebler W., Dierich M.P., Weiss E.H.;  
RT "Cloning of the 1.4-kb mRNA species of human complement factor H  
reveals a novel member of the short consensus repeat family related  
to the carboxy terminal of the classical 150-kDa molecule.";  
RL J. Immunol. 146:3190-3196(1991).  
RN [5]  
RP SEQUENCE OF 19-35.  
RX MEDLINE: 83048213.  
RA Sim R.B., Discipio R.G.;  
RT "Purification and structural studies on the complement-system control  
protein beta 1H (Factor H).";  
RL Biochem. J. 205:285-293(1982).  
RN [6]  
RP STRUCTURE BY NMR OF 927-985 (SUSHI 16).  
RX MEDLINE: 91278057.  
RA Norman D.G., Barlow P.N., Baron M., Day A.J., Sim R.B., Campbell I.D.;  
RT "Three-dimensional structure of a complement control protein module  
in solution.";  
RL J. Mol. Biol. 219:717-725(1991).

```

RN [7]
RP STRUCTURE BY NMR OF 264-322 (SUSHI 5).
RX MEDLINE; 92232649.
RA Barlow P.N., Norman D.G., Steinkasserer A., Horne T.J., Pearce J.,
RT Discolli P.C., Sim B., Campbell I.D.;
RA "Solution structure of the fifth repeat of factor H: a second example
RT of the complement control protein module.",
RL Biochemistry 31:3626-3634(1992).
RN [8]
RP STRUCTURE BY NMR OF 866-985 (SUSHIS 15 AND 16).
RX MEDLINE; 93323119.
RA Barlow P.N., Steinkasserer A., Norman D.G., Kieffer B., Wiles A.P.,
RT "Solution structure of a pair of complement modules by nuclear
RT magnetic resonance.",
RL J. Mol. Biol. 232:268-284(1993).
CC -1- FUNCTION: FACTOR H FUNCTIONS AS A COFACTOR IN THE INACTIVATION OF
CC C3B BY FACTOR I AND ALSO INCREASES THE RATE OF DISSOCIATION OF THE
CC C3BBB COMPLEX (C3 CONVERTASE) AND THE (C3B)NBB COMPLEX (C5
CC CONVERTASE) IN THE ALTERNATIVE COMPLEMENT PATHWAY.
CC -1- SIMILARITY: CONTAINS 20 SUSHI (SCR) REPEATS.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION
CC 341 ONWARD DUE TO A FRAMESHIFT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y00716; CAA68704.1; -
DR EMBL; X04697; CAA81739.1; ALT_FRAME.
DR EMBL; M65294; AAA35948.1; -
DR PIR; S00254; NBHHD.
DR PIR; S03013; S03013.
DR PDB; 1HCC; 15-APR-92.
DR PDB; 1HHI; 15-JUL-93.
DR MIM; 134370; -
DR INTERPRO: IPR000436; -
DR PFAM; PF00084; sushi; 20.
KW Complement alternate pathway; Plasma; Glycoprotein; Repeat; Sushi;
KW Signal; 3D-structure; Polymorphism.
FT SIGNAL 1 18
FT CHAIN 19 1231
FT DOMAIN 20 1229
FT REPEAT 20 81
FT REPEAT 84 142
FT REPEAT 145 206
FT REPEAT 209 263
FT REPEAT 324 386
FT REPEAT 388 443
FT REPEAT 447 506
FT REPEAT 508 565
FT REPEAT 568 624
FT REPEAT 629 685
FT REPEAT 745 804
FT REPEAT 810 865
FT REPEAT 869 927
FT REPEAT 930 985
FT REPEAT 988 1044
FT REPEAT 1047 1103
FT REPEAT 1108 1164
FT REPEAT 1166 1229
FT DISULFID 21 66
FT DISULFID 32 80
FT DISULFID 85 129
FT DISULFID 114 141

```

```

FT DISULFID 146 192
FT DISULFID 178 205
FT DISULFID 210 251
FT DISULFID 237 262
FT DISULFID 267 309
FT DISULFID 294 320
FT DISULFID 325 374
FT DISULFID 357 385
FT DISULFID 389 431
FT DISULFID 416 442
FT DISULFID 448 494
FT DISULFID 477 505
FT DISULFID 509 553
FT DISULFID 536 564
FT DISULFID 569 611
FT DISULFID 597 623
FT DISULFID 630 673
FT DISULFID 659 684
FT DISULFID 691 733
FT DISULFID 719 744
FT DISULFID 753 792
FT DISULFID 781 803
FT DISULFID 811 853
FT DISULFID 839 864
FT DISULFID 870 915
FT DISULFID 901 926
FT DISULFID 931 973
FT DISULFID 959 984
FT DISULFID 989 1032
FT DISULFID 1018 1043
FT DISULFID 1048 1091
FT DISULFID 1077 1102
FT DISULFID 1109 1152
FT DISULFID 1138 1163
FT DISULFID 1167 1218
FT DISULFID 1201 1228
FT CARBOHYD 529 529
FT CARBOHYD 718 718
FT CARBOHYD 802 802
FT CARBOHYD 822 822
FT CARBOHYD 882 882
FT CARBOHYD 911 911
FT CARBOHYD 1029 1029
FT CARBOHYD 1095 1095
FT VARIANT 402 402
FT CONFLICT 21 21
FT CONFLICT 30 30
FT CONFLICT 34 34
FT CONFLICT 53 54
FT STRAND 870 871
FT STRAND 876 876
FT TURN 877 878
FT STRAND 879 881
FT STRAND 890 891
FT TURN 893 894
FT STRAND 896 900
FT TURN 903 904
FT STRAND 907 907
FT STRAND 912 916
FT TURN 917 918
FT STRAND 919 920
FT STRAND 926 926
SQ SEQUENCE 1231 AA; 139125 MW; C65EC8CF8600B3FD CRC64;

```

```

Query Match 100.0%; Score 1497; DB 1; Length 1231;
Best Local Similarity 100.0%; Pred. No. 1,2e-114;
Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 EDCNLPFRNRTEITLTSMSDQTYEGTQAIYKCPGRSGNVIWCRKGWALNPLR 60
Db 19 EDCNELPFRNRTEITLTSMSDQTYEGTQAIYKCPGRSGNVIWCRKGWALNPLR 78

```





```

FT SIGNAL 1 48
FT CHAIN 49 597 C4B-BINDING PROTEIN ALPHA CHAIN.
FT DOMAIN 49 539 8 X SUSHI (SCR) REPEATS.
FT REPEAT 49 109 SUSHI 1.
FT REPEAT 112 171 SUSHI 2.
FT REPEAT 174 235 SUSHI 3.
FT REPEAT 238 295 SUSHI 4.
FT REPEAT 298 361 SUSHI 5.
FT REPEAT 364 423 SUSHI 6.
FT REPEAT 425 481 SUSHI 7.
FT REPEAT 483 539 SUSHI 8.
FT DISULFID 50 96 BY SIMILARITY.
FT DISULFID 81 108 BY SIMILARITY.
FT DISULFID 113 154 BY SIMILARITY.
FT DISULFID 140 170 BY SIMILARITY.
FT DISULFID 175 217 BY SIMILARITY.
FT DISULFID 203 234 BY SIMILARITY.
FT DISULFID 239 281 BY SIMILARITY.
FT DISULFID 267 294 BY SIMILARITY.
FT DISULFID 329 348 BY SIMILARITY.
FT DISULFID 332 360 BY SIMILARITY.
FT DISULFID 365 387 BY SIMILARITY.
FT DISULFID 426 468 BY SIMILARITY.
FT DISULFID 454 480 BY SIMILARITY.
FT DISULFID 484 525 BY SIMILARITY.
FT DISULFID 511 538 BY SIMILARITY.
FT DISULFID 546 556 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT DISULFID 558 558 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT CARBOHYD 221 221 N-LINKED (GLCNAC. . .).
FT CARBOHYD 506 506 N-LINKED (GLCNAC. . .).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .).
FT VARIANT 92 92 Q -> T.
FT VARIANT 357 357 /FTIG-VAR_001977.
FT VARIANT 357 357 Y -> H.
FT VARIANT 357 357 /FTIG-VAR_001978.
SQ SEQUENCE 597 AA; 67033 MW; 67E03F2EA85A16DD CRC64;

```

Query Match 20.2%; Score 303; DB 1; Length 597;  
 Best Local Similarity 30.7%; Pred. No. 1.5e-17;  
 Matches 80; Conservative 42; Mismatches 101; Indels 38; Gaps 16;

```

QY 20 SDQTYEGTQAIYKCRPGY-RSLGNVIMWCRK-GEWVALNPLRCKCRPGCHGPTPECT 77
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 67 TETRFKGTGLTKYTCLPYGVRSHTQTLCNSDGEWV-YWTF-CIYKCRHHPG----- 118
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 78 FTLTGAVN-----FEYGVKAVYTCNGSYOLGEINRECDTD---GWNDDIPICEVVKC 128
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 119 --LRNGVEIKTDLSPGSGSEGFLLIGSTSR-CEVQDRGWSHPLPQCEIVKC 175
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 129 LPVYAPENKIVSSAMPEDEHYHFGQAVRFVCSGYKIEGDEMHGSDS---GFWSKER 184
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 176 KPPEDINGNHSGE---ENFVAYGVSIVYSCDPRSLIGHASISCTVEVETIGVWRPSP 221
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 185 PKCVETSCSPDYVINGSPISQ-KIYKENRFQYCKCMGYEYERGDV--C-TEGQWRP 240
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 232 PTCEKIKICRRPDVSHGEMVSGFSPYVYKDTIVFKCCKGFLV--RGSSTICHADSKMWP 289
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 241 L-PSCEKSCDN-PYIPNGDY 259
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 290 SPSPACEPNSCINLPDIPHASW 310
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 4
VCP_VACCV STANDARD; PRT; 263 AA.
AC P10998;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE COMPLEMENT CONTROL PROTEIN PRECURSOR (VCP) (SECRETORY PROTEIN 35)

```

```

DE (PROTEIN C3) (28 KDA PROTEIN).
GN C3.
OS Vaccinia virus (strain WR), and Vaccinia virus (strain Copenhagen).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 20-37.
RC STRAIN-WR.
RX MEDLINE; 88318974.
RA Kotwal G.J., Moss B.;
RT "Vaccinia virus encodes a secretory polypeptide structurally related
to complement control proteins.";
RL Nature 335:176-178(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-WR.
RX MEDLINE; 89073756.
RA Kotwal G.J., Moss B.;
RT "Analysis of a large cluster of nonessential genes deleted from a
vaccinia virus terminal transposition mutant.";
RL Virology 167:524-537(1988).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-COPENHAGEN;
RX MEDLINE; 91021027.
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RT "The complete DNA sequence of vaccinia virus.";
RL Virology 179:247-266(1990).
RN [4]
RP COMPLETE GENOME.
RC STRAIN-COPENHAGEN;
RX Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RT Paolelli E.;
RL Virology 179:517-563(1990).
RN [5]
RP FUNCTION.
RX MEDLINE; 92115714.
RA Isaacs S.N., Kotwal G.J., Moss B.;
RT "Vaccinia virus complement-control protein prevents
antibody-dependent complement-enhanced neutralization of infectivity
and contributes to virulence.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:628-632(1992).
RN [6]
RP STRUCTURE BY NMR OF 146-263.
RX MEDLINE; 97446168.
RA Wiles A.P., Shaw G., Bright J., Perczel A., Campbell I.D.,
RT Barlow P.N.;
RL "NMR studies of a viral protein that mimics the regulators of
complement activation.";
RN [7]
RP FUNCTION: SERVES TO PROTECT THE VIRUS AGAINST COMPLEMENT ATTACK BY
INHIBITING BOTH CLASSICAL AND ALTERNATIVE PATHWAYS OF COMPLEMENT
ACTIVATION. BINDS C3B AND C4B.
RN [8]
RP SIMILARITY: BELONGS TO THE SUPERFAMILY OF THE REGULATORS OF
COMPLEMENT ACTIVATION (RCA).
RN [9]
RP SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS.
RN [10]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X13166; CAA31564.1; -.
CC EMBL; M22812; AAA69605.1; -.
CC EMBL; M35027; AAA47997.1; -.
CC PIR; A31005; MWV2SP.
CC PDB; 1YVC; 03-DEC-97.
CC PDB; 1YVD; 03-DEC-97.
CC PDB; 1YVE; 03-DEC-97.

```

```

DR INTERPRO: IPR000436;
DR PFAM: PF00084; sushi; 4
KW Signal; Repeat; Sushi; 3d-structure.
FT SIGNAL 1 19
FT CHAIN 20 263
FT DOMAIN 20 263 COMPLEMENT CONTROL PROTEIN.
FT REPEAT 20 82 4 X SUSHI (SCR) REPEATS.
FT REPEAT 85 144 SUSHI 2.
FT REPEAT 147 202 SUSHI 3.
FT REPEAT 205 262 SUSHI 4.
FT DISULFID 21 70 BY SIMILARITY.
FT DISULFID 54 81 BY SIMILARITY.
FT DISULFID 86 126 BY SIMILARITY.
FT DISULFID 112 143 BY SIMILARITY.
FT DISULFID 148 190 BY SIMILARITY.
FT DISULFID 176 201 BY SIMILARITY.
FT DISULFID 206 248 BY SIMILARITY.
FT DISULFID 234 261 BY SIMILARITY.
SQ SEQUENCE 263 AA; 28629 MW; EA322C9A6EF8997 CRC64;

Query Match 20.1%; Score 301.5; DB 1; Length 263;
Best Local Similarity 33.3%; Pred. No. 7.9e-18;
Matches 78; Conservative 29; Mismatches 106; Indels 21; Gaps 12;

QY 20 SDQYPEGQAIYKGRPGYR--SIGNVIMVCKGKGVVALNPLKCKQKRCPCGHPTPGT 77
DB 40 ANANINIDTLEYLCLPGRKCKMGPIYAKCTGTWLPN--QCIRKRCPSPRDIDNQ 96
QY 78 FTLTGAVFVEYGVKAVYTCNKGYSLLIGET-NYRECDDG--WTNDIPICEVWKLPTA 133
DB 97 LDIGG---VDFGSSITYSNCGSHLIGESKSCYCELGSGMWNNEAPRCEVCKQSPS 153
QY 134 PENGKIVSSAEPREHYFGQAVRVCNSGKIEDEMHCSDDGFWKSKPKVEISCK 193
DB 154 ISNGR--HNGYED--FYTDGVSIVYSCNSGYSLGNSGLCS-GGEWS-DPPTCQIVKCP 207
QY 194 SPDVINGSPISQ-KIYKENERFOYKCMNGEYSEPDGAVCTESG-WRP-LPSC 244
DB 208 HPTISNGYLSGFKRYSYNONVDEKCKYKYLSSSSSTCSFGNTWAKPELPKC 261

RESULT 5
ID CRI_HUMAN STANDARD; PRT; 2039 AA.
AC P17927;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE COMPLEMENT RECEPTOR TYPE 1 PRECURSOR (C3B/C4B RECEPTOR) (CD35
DE ANTIGEN).
GN CRI OR C3BR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87168191.
RA Klickstein L.B., Wong W.W., Smith J.A., Weis J.H., Wilson J.G.,
RA Fearon D.T.,
RT "Human C3b/C4b receptor (CRI). Demonstration of long homologous
RT repeating domains that are composed of the short consensus repeats
RT characteristics of C3/C4 binding proteins."
RL J. Exp. Med. 165:1095-1112(1987).
RN [3]

```

```

RP SEQUENCE OF 761-783; 831-845 AND 1179-1195 FROM N.A.
RX MEDLINE; 86067975.
RA Wong W.W., Klickstein L.B., Smith J.A., Weis J.H., Fearon D.T.;
RT "Identification of a partial cDNA clone for the human receptor for
RT complement fragments C3b/C4b."
RL Proc. Natl. Acad. Sci. U.S.A. 82:7711-7715(1985).
CC -1- FUNCTION: CRI, PRESENT ON ERYTHROCYTES, LEUCOCYTES, GLOMERULAR
CC PODOCYTES, AND SPLENIC FOLLICULAR DENDRITIC CELLS, MEDIATES THE
CC BINDING BY THESE CELLS OF PARTICLES AND IMMUNE COMPLEXES THAT HAVE
CC ACTIVATED COMPLEMENT.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- POLYMORPHISM: CRI IS RESPONSIBLE FOR THE KNOBS BLOOD GROUP SYSTEM.
CC -1- MISCELLANEOUS: SEVEN SHORT CONSENSUS REPEATS (SCR) CONSTITUTE A
CC LONG HOMOLOGOUS REPEAT (LHR). THE N-TERMINAL TWO SRCS OF LHR-A
CC CONTAINED A SITE DETERMINING C4 SPECIFICITY, AND THE N-TERMINAL
CC TWO SRCS OF LHR-B AND -C EACH HAD A SITE DETERMINING C3
CC SPECIFICITY.
CC -1- MISCELLANEOUS: THIS IS THE SEQUENCE OF THE F ALLOTYPE OF CRI.
CC -1- SIMILARITY: CONTAINS 30 SUSHI (SCR) REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M11569; AAA52297.1; -
DR EMBL; M11617; AAA52298.1; -
DR EMBL; M11618; AAA52299.1; -
DR EMBL; Y00816; CAA68755.1; -
DR EMBL; X05309; CAA28933.1; -
DR PIR; A28507; A28507-
DR PIR; A24748; A24748-
DR PIR; B24748; B24748-
DR PIR; C24748; C24748-
DR PIR; S03843; S03843-
DR HSSP; P08603; 1HFI.
DR MIM; 120620; -
DR INTERPRO: IPR000436; -
DR PFAM: PF00084; sushi; 30.
KW Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
KW Receptor; Sushi; Blood group antigen.
FT SIGNAL 1 41
FT CHAIN 2 2039
FT DOMAIN 42 1971
FT TRANSMEM 1972 1996
FT DOMAIN 1997 2039
FT MOD_RES 42 42
FT DOMAIN 42 488
FT REPEAT 42 100
FT REPEAT 103 162
FT REPEAT 165 233
FT REPEAT 237 294
FT REPEAT 296 354
FT REPEAT 357 417
FT REPEAT 420 488
FT DOMAIN 492 938
FT REPEAT 492 550
FT REPEAT 553 612
FT REPEAT 615 683
FT REPEAT 687 744
FT REPEAT 746 804
FT REPEAT 807 867
FT REPEAT 870 938
FT DOMAIN 942 1388
FT REPEAT 942 1000
FT REPEAT 1003 1062
CC COMPLEMENT RECEPTOR TYPE 1.
CC EXTRACELLULAR (POTENTIAL).
CC POTENTIAL.
CC CYTOPLASMIC (POTENTIAL).
CC PYRROLIDONE CARBOXYLIC ACID (POTENTIAL).
CC 7 X SUSHI (SCR) REPEATS, LONG HOMOLOGOUS
CC REPEAT A.
CC SUSHI A1.
CC SUSHI A2.
CC SUSHI A3.
CC SUSHI A4.
CC SUSHI A5.
CC SUSHI A6.
CC SUSHI A7.
CC 7 X SUSHI
CC REPEAT B.
CC SUSHI B1.
CC SUSHI B2.
CC SUSHI B3.
CC SUSHI B4.
CC SUSHI B5.
CC SUSHI B6.
CC SUSHI B7.
CC 7 X SUSHI
CC REPEAT C.
CC SUSHI C1.
CC SUSHI C2.

```

```

FT REPEAT 1065 1133 SUSHI C3.
FT REPEAT 1137 1194 SUSHI C4.
FT REPEAT 1196 1254 SUSHI C5.
FT REPEAT 1257 1317 SUSHI C6.
FT REPEAT 1320 1388 SUSHI C7.
FT DOMAIN 1395 1486 7 X SUSHI (SCR) REPEATS, LONG HOMOLOGOUS
FT REPEAT 1395 1453 REPEAT D.
FT REPEAT 1456 1515 SUSHI D1.
FT REPEAT 1518 1586 SUSHI D2.
FT REPEAT 1590 1647 SUSHI D3.
FT REPEAT 1649 1707 SUSHI D4.
FT REPEAT 1710 1770 SUSHI D5.
FT REPEAT 1773 1841 SUSHI D6.
FT REPEAT 1847 1905 SUSHI D7.
FT REPEAT 1908 1966 2 X SUSHI (SCR) REPEATS.
FT REPEAT 1908 1966 SUSHI 1.
FT DISULFID 43 86 BY SIMILARITY.
FT DISULFID 73 99 BY SIMILARITY.
FT DISULFID 104 145 BY SIMILARITY.
FT DISULFID 131 161 BY SIMILARITY.
FT DISULFID 166 215 BY SIMILARITY.
FT DISULFID 195 232 BY SIMILARITY.
FT DISULFID 238 280 BY SIMILARITY.
FT DISULFID 266 293 BY SIMILARITY.
FT DISULFID 297 340 BY SIMILARITY.
FT DISULFID 326 353 BY SIMILARITY.
FT DISULFID 358 400 BY SIMILARITY.
FT DISULFID 366 416 BY SIMILARITY.
FT DISULFID 421 470 BY SIMILARITY.
FT DISULFID 450 536 BY SIMILARITY.
FT DISULFID 493 549 BY SIMILARITY.
FT DISULFID 523 595 BY SIMILARITY.
FT DISULFID 554 611 BY SIMILARITY.
FT DISULFID 581 665 BY SIMILARITY.
FT DISULFID 616 682 BY SIMILARITY.
FT DISULFID 645 730 BY SIMILARITY.
FT DISULFID 688 743 BY SIMILARITY.
FT DISULFID 716 790 BY SIMILARITY.
FT DISULFID 747 803 BY SIMILARITY.
FT DISULFID 776 850 BY SIMILARITY.
FT DISULFID 808 866 BY SIMILARITY.
FT DISULFID 836 920 BY SIMILARITY.
FT DISULFID 871 937 BY SIMILARITY.
FT DISULFID 900 986 BY SIMILARITY.
FT DISULFID 943 999 BY SIMILARITY.
FT DISULFID 973 1045 BY SIMILARITY.
FT DISULFID 1004 1061 BY SIMILARITY.
FT DISULFID 1031 1115 BY SIMILARITY.
FT DISULFID 1066 1132 BY SIMILARITY.
FT DISULFID 1095 1180 BY SIMILARITY.
FT DISULFID 1138 1193 BY SIMILARITY.
FT DISULFID 1166 1240 BY SIMILARITY.
FT DISULFID 1197 1253 BY SIMILARITY.
FT DISULFID 1226 1300 BY SIMILARITY.
FT DISULFID 1258 1316 BY SIMILARITY.
FT DISULFID 1286 1370 BY SIMILARITY.
FT DISULFID 1321 1387 BY SIMILARITY.
FT DISULFID 1350 1439 BY SIMILARITY.
FT DISULFID 1396 1452 BY SIMILARITY.
FT DISULFID 1426 1498 BY SIMILARITY.
FT DISULFID 1457 1514 BY SIMILARITY.
FT DISULFID 1484 1568 BY SIMILARITY.
FT DISULFID 1519 1585 BY SIMILARITY.
FT DISULFID 1548 1633 BY SIMILARITY.
FT DISULFID 1591 1646 BY SIMILARITY.
FT DISULFID 1619 1693 BY SIMILARITY.
FT DISULFID 1650 1706 BY SIMILARITY.
FT DISULFID 1679 1753 BY SIMILARITY.
FT DISULFID 1711 1769 BY SIMILARITY.
FT DISULFID 1731 1823 BY SIMILARITY.
FT DISULFID 1774 1840 BY SIMILARITY.
FT DISULFID 1803 1840 BY SIMILARITY.

```

```

FT DISULFID 1848 1891 BY SIMILARITY.
FT DISULFID 1877 1904 BY SIMILARITY.
FT DISULFID 1909 1952 BY SIMILARITY.
FT DISULFID 1938 1965 BY SIMILARITY.
FT CARBOHYD 56 56 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 252 252 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 447 447 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 509 509 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 702 702 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 860 860 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 897 897 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 959 959 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1028 1028 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1152 1152 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1310 1310 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1481 1481 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1504 1504 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1534 1534 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1540 1540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1605 1605 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1763 1763 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1908 1908 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 2039 AA; 223589 MW; B2FD29B6AD3C5EB7 CRC64;

Query Match 19.9%; Score 298.5; DB 1; Length 2039;
Best Local Similarity 27.0%; Pred. No. 1.4e-16;
Matches 86; Conservative 49; Mismatches 111; Indels 73; Gaps 18;

OY 3 CNE--LPRRRTFELTGSWSIDQYPTQTAIYKCRPGYSRLGNVYVCRKGEWVALNPL 59
DB 43 CNAPEMLPAPARTNLT-----DEFEPPICTIYLYNECRPGYSRPSILTKMSVTMGARD- 97
OY 60 KRCQKRPCHGEDPFPFGFTLTGCVNFEYGYKAVYTCNEGQVLGEINRYRC-----DTDG 115
DB 98 -RCRRKSCNPPDPVNGWVHYTKG--IQGSGIKYSCRYKGLIGS--SSATCIISGDVI 153
OY 116 WTNDDPICCVNVC-LPVTAPENGKIVSSAMEPREDEYHFGQAVRFVCSNGS-----YKIEG 168
DB 154 WDNETPICRIPCGILPPT-TNGDPISITNRE---NFHYGSVYTYCNGSSGKRYFELVG 209
OY 169 DEEMHC--SDD--GFWSKKPKC-VEISCKSPDVINGSPIS-OKIIVENERFOYKCNMG 222
DB 210 EPSICTSDDDQVIGMSGAPQCIIIPNKCTPPVNGILVSDNRLSLFNEVEFRCPG 269
OY 223 YEYSEKGAIVCTE-SQWR-LPSCBE-----KSCDNP 252
DB 270 FVMKGPFRVYKCOLKKNWELPSCSRVCOPPDVLAERTORDKDNFSPGQEVFYSCPG 329
OY 253 Y-----IPNGDYSP 261
DB 330 YDLRGAASMRCTPQGDWSP 348

RESULT 6
CABP RAT STANDARD; PRT; 558 AA.
AC 063514;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE C4B-BINDING PROTEIN ALPHA CHAIN PRECURSOR.
GN C4BPA.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NC [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-LIVER;
RX MEDLINE; 97166082.
RA Hillarp A., Wiklund H., Thern A., Dahlback B.;
RT "Molecular cloning of rat C4b binding protein alpha- and beta-chains:

```









FT	CARBOHYD	376	376	N-LINKED (GLCNAC...)	(POTENTIAL)
FT	CARBOHYD	400	400	N-LINKED (GLCNAC...) <th>(POTENTIAL)</th>	(POTENTIAL)
FT	CONFLICT	253	253	C -> Y (IN REF. 2).	
FT	CONFLICT	313	313	L -> F (IN REF. 2).	
FT	CONFLICT	321	321	T -> N (IN REF. 2).	
FT	CONFLICT	327	327	K -> N (IN REF. 2).	
FT	CONFLICT	363	363	V -> A (IN REF. 2).	
FT	CONFLICT	384	384	V -> M (IN REF. 2).	
FT	CONFLICT	461	484	KVPSSSECSICQDNGNSQWPSLI ->	
FT	CONFLICT	461	484	NIFLPAAPRAEYPMDFKCLTL (IN REF. 2).	
FT	CONFLICT	461	484	AFEF74FE25C1FD013 CRC64;	
SO	SEQUENCE	484 AA;	52567 MW;	AFEF74FE25C1FD013 CRC64;	
Query Match					
Best Local Similarity		17.8%;	Score 266;	DB 1;	Length 484;
Matches		65;	Conservative	42;	Mismatches 103;
					Indels 30;
					Gaps 12;
QY	22 QYRPETQALTYKCRPRYBSLGNVIMVC-RKGEWALNPLRKCKRRC--GHGDTPTFCT	77			
DB	196 QSLPMWTTCAFECKKEEFELIGREHLQCTSSGSDWKRP--TKKAVYCDYVGHPON---GD	250			
QY	78 FTLLTGAVNEEYCYKAV--YTCNMGYQLLGEINVRBEDPDG--FTNIDIPICEYVKCLPYTAP	134			
DB	251 VACNHSISIEFAKSGICHTCAAGGELGSPADP-ECTAOGQMTQAPYCAKCAVAVSP	309			
QY	135 ENGKIVSSAMBPDRREYHFGQAVREVCNSGKYTEGDEMHCSDDGFWSKRKPVCVEISKS	194			
DB	310 KNG-LVKEFTHSPTEGETFYKSCAFCEEGFELRGSQAOLACTSQGQMTQDEVPSCQYVQCSS	368			
QY	195 PVV-----INGSPISQKIIYKENRPOYKCKNMKEYSERDVAICTSG-WR-PLPSC	245			
DB	369 LEVPELNNSCSGEYFVAVC-----TFACPDEGMNLGVALTGATGHWGMLPCE	421			
RESULT 11					
CABP_BOVIN					
ID	CABP_BOVIN	STANDARD;	PRT;	610 AA.	
AC	028065;				
DT	01-NOV-1997 (Rel. 35, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	01-NOV-1997 (Rel. 35, Last annotation update)				
DE	C4B-BINDING PROTEIN ALPHA CHAIN PRECURSOR.				
GN	C4BPA.				
OS	Bos taurus (Bovine).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;				
OC	Bovidae; Bovinae; Bos.				
OC	[1]				
RN	SEQUENCE FROM N.A.				
RP	TISSUE=LIVER;				
RX	MEDLINE: 95015909.				
RA	Hillarp A., Thern A., Dahlbaech B.;				
RT	"Bovine C4b binding protein. Molecular cloning of the alpha- and				
RT	beta-chains provides structural background for lack of complex				
RT	formation with protein S.";				
RL	J. Immunol. 153:4190-4199(1994).				
CC	-1- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT				
CC	ACTIVATION. IT BINDS AS A COPACOPOR TO C3B/C4B INACTIVATOR				
CC	(C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT				
CC	ALSO ACCELERATES THE DEGRADATION OF THE C4B2A COMPLEX (C3				
CC	CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA				
CC	CHAIN BINDS C4B. IT INTERACTS ALSO WITH SERUM A2M AND P COMPONENT.				
CC	-1- FUNCTION: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS				
CC	(BY SIMILARITY).				
CC	-1- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.				
CC	-1- SIMILARITY: CONTAINS 8 SUSHI (SCR) REPEATS.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> )				

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

CC EMBL: Z31693; CAA83498.1; -.  
 DR HSSP; P10998; IVC.  
 DR INTERPRO: IPR000436; -.  
 DR PFAM: PF00084; sushi; 8.  
 KW Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal.  
 FT SIGNAL 1 48  
 FT CHAIN 49 610  
 FT DOMAIN 49 542  
 FT REPEAT 49 108  
 FT REPEAT 111 170  
 FT REPEAT 173 235  
 FT REPEAT 238 295  
 FT REPEAT 298 363  
 FT REPEAT 366 426  
 FT REPEAT 428 484  
 FT REPEAT 486 542  
 FT DISULFID 50 95  
 FT DISULFID 80 107  
 FT DISULFID 112 153  
 FT DISULFID 139 169  
 FT DISULFID 174 217  
 FT DISULFID 203 234  
 FT DISULFID 239 281  
 FT DISULFID 267 294  
 FT DISULFID 299 350  
 FT DISULFID 334 362  
 FT DISULFID 7366 7390  
 FT DISULFID 7367 7412  
 FT DISULFID 7402 425  
 FT DISULFID 429 471  
 FT DISULFID 457 483  
 FT DISULFID 487 528  
 FT DISULFID 514 541  
 FT DISULFID 549 561  
 FT DISULFID 561 561  
 FT CARBOHYD 66 66  
 FT CARBOHYD 221 221  
 FT CARBOHYD 525 525  
 FT CARBOHYD 602 602  
 SQ SEQUENCE 610 AA; 68886 MW; D806B270E8A06B58 CRC64;  
 Query Match 17.7%; Score 265.5; DB 1; Length 610;  
 Best Local Similarity 27.5%; Pred. No. 1.8e-14;  
 Matches 76; Conservative 45; Mismatches 110; Indels 45; Gaps 16;  
 Oy 6 LPPRRNEILLGSMDSQTPGCTQAIYKRCGYR--SIGNVIMCRKGEWALNLRKQ 63  
 Db 52 IPPYLDRAFPINELNRETFETGTLRYCRPGYRISKRNLICDGTDM--KYKECV 108  
 Oy 64 KRPGCHGDPFPGFTFLGCVN-----FEYGVKAVYTCNEGQYLGEIN-YRECDTDS-- 115  
 Db 109 KKRENGE-----LLNGQVIYKTDSPFSEIEFSCSEGVILGANSYCOQLODKV 161  
 Oy 116 WTNDIPICEVVKLPVAPANGKIVSSAMEPDREYHGSQAVRPNCSYKKEGDEMHCS 175  
 Db 162 WSDPLPOCIILAKCEPPPTISNGR--HNGDDED-FYTYGSSVYSCDRDFSMIGKASISCR 218  
 Oy 176 DD-----GFWSEKPKVCVIEISCKSPDIVINGSPIISKI-----IYKENRFPQYKCMGEX 225  
 Db 219 VENKTIGWSSPPSCCKVICQVQPVYKDG-----KITSGFPGIYTYQOSIYACKNGFRL 273  
 Oy 226 SERGDV--C-TESGWR-PLPSCSEKSC-DNPIYIPN 256  
 Db 274 --EGDSLHCADNSMWNPPPTCELNGCLGLPHIPH 307  
 RESULT 12  
 CCPH\_HSVSA STANDARD; PRT; 360 AA.  
 ID\_CCPH\_HSVSA  
 AC 001016;

DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE COMPLEMENT CONTROL PROTEIN HOMOLOG PRECURSOR (CCPH).  
 GN 4 OR CCPH.  
 OS Herpesvirus saimiri (strain 11).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Gammaherpesvirinae; Rhadinovirus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92333688.  
 RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,  
 RA Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B.,  
 RA Honess R.W.,  
 RT "Primary structure of the herpesvirus saimiri genome.";  
 RL J. Virol. 66:5047-5058(1992).  
 RN [2]  
 RP SIMILARITY TO CCP.  
 RX MEDLINE: 92260674.  
 RA Albrecht J.-C., Fleckenstein B.,  
 RT "New member of the multigene family of complement control proteins in  
 RT herpesvirus saimiri.";  
 RL J. Virol. 66:3937-3940(1992).  
 CC -I- ALTERNATIVE PRODUCTS: A MEMBRANE-BOUND FORM AND A SECRETED FORM  
 CC ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.  
 CC -I- SIMILARITY: BELONGS TO THE SUPERFAMILY OF THE REGULATORS OF  
 CC COMPLEMENT ACTIVATION (RCA).  
 CC -I- SIMILARITY: CONTAINS 3 SUSHI (SCR) REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X64346; CAA45626.1; -.  
 DR EMBL: X64346; CAA45627.1; -.  
 DR EMBL: X60283; CAA42823.1; -.  
 DR EMBL: X60283; CAA42822.1; -.  
 DR PIR: B42534; WMBE2E.  
 DR PIR: A42534; WMBE1E.  
 DR PIR: S24567; S24567.  
 DR HSSP; P10998; IVC.  
 DR INTERPRO: IPR000436; -.  
 DR PFAM: PF00084; sushi; 4.  
 KW Signal; Repeat; Sushi; Transmembrane; Alternative splicing;  
 KW Glycoprotein.  
 FT SIGNAL 1 20  
 FT CHAIN 21 360  
 FT DOMAIN 83 265  
 FT REPEAT 83 143  
 FT REPEAT 146 206  
 FT REPEAT 209 265  
 FT TRANSMEM 328 350  
 FT DISULFID 84 125  
 FT DISULFID 111 142  
 FT DISULFID 147 191  
 FT DISULFID 175 205  
 FT DISULFID 210 252  
 FT DISULFID 238 264  
 FT CARBOHYD 36 36  
 FT CARBOHYD 39 39  
 FT CARBOHYD 46 46  
 FT CARBOHYD 72 72  
 FT CARBOHYD 155 155  
 FT CARBOHYD 294 294  
 FT CARBOHYD 289 302  
 FT VARSPLIC 303 360  
 FT SEQUENCE 360 AA; 40006 MW; 6278A6C2ECD49669 CRC64;  
 POTENTIAL.  
 COMPLEMENT CONTROL PROTEIN HOMOLOG.  
 3 X SUSHI (SCR) REPEATS.  
 SUSHI 1.  
 SUSHI 2.  
 SUSHI 3.  
 POTENTIAL.  
 BY SIMILARITY.  
 BY SIMILARITY.  
 BY SIMILARITY.  
 BY SIMILARITY.  
 BY SIMILARITY.  
 BY SIMILARITY.  
 BY SIMILARITY.  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 RINGNCTTSMPTQ -> AECACPGSNTPISS (IN  
 SHORT ISOFORM).  
 MISSING (IN SHORT ISOFORM).  
 6278A6C2ECD49669 CRC64;



```

CC or send an email to license@sib.ch).
CC -----
DR EMBL: M31516; AAA52169.1; -
DR EMBL: M30142; AAA52168.1; -
DR EMBL: M15799; AAA52167.1; -
DR EMBL: M64653; AAA52170.1; -
DR EMBL: M64356; AAA52170.1; JOINED.
DR EMBL: S72858; AAC60633.1; -
DR PIR: B26359; B26359.
DR PIR: A26359; A26359.
DR PIR: S16187; S16187.
DR PIR: A39101; A39101.
DR PIR: S23138; S23138.
DR HSSP: P08603; IHCC.
DR MIM: 125240; -.
DR INTERPRO: IPR000436; -.
DR PFAM: PF00084; sushi; 4.
KW Complement pathway; Plasma; Glycoprotein; Membrane; Repeat;
KW Alternative splicing; GPI-anchor; Signal; Sushi; Polymorphism;
KW Blood group antigen.
FT SIGNAL 1 34
FT CHAIN 35 353 COMPLEMENT DECAT-ACCELERATING FACTOR.
FT PROPEP 354 381 REMOVED IN MATURE FORM.
FT DOMAIN 35 284 4 X SUSHI (SCR) REPEATS.
FT REPEAT 35 95 SUSHI 1.
FT REPEAT 97 159 SUSHI 2.
FT REPEAT 162 221 SUSHI 3.
FT REPEAT 224 284 SUSHI 4.
FT DOMAIN 287 356 SER/THR-RTCH.
FT DISULFID 36 81
FT DISULFID 65 94
FT DISULFID 98 145
FT DISULFID 129 158
FT DISULFID 163 204
FT DISULFID 190 220
FT DISULFID 225 267
FT DISULFID 253 283
FT CARBOHYD 95 95
FT LIPID 353 353
FT VARSPPLIC 362 381
FT FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT GPI-ANCHOR.
FT FT HTCTLTGLGLTGLTGLTGLT -> SRPVTQAGNRWCNSSL
FT FT OSRTPEKRSFHFSLPSSWYRAHVHDFVAFADASNHGIA
FT FT DIAKEELRRKYTGVRFLFVS (IN ISOFORM DAF-1).
FT FT R -> L (IN TC(B) ANTIGEN).
FT FT /FTID-VAR_001997.
FT FT R -> P (IN TC(C) ANTIGEN).
FT FT /FTID-VAR_001998.
FT FT L -> R (IN MES(A) ANTIGEN).
FT FT /FTID-VAR_001999.
FT FT S -> L (IN DR(A-) ANTIGEN).
FT FT /FTID-VAR_002000.
FT FT /FTID-VAR_002001.
FT FT A -> P (IN CR(A-) ANTIGEN).
FT FT /FTID-VAR_002001.
FT FT T -> I (IN REF. 1).
FT FT S -> M (IN REF. 2).
FT FT S -> M (IN REF. 2).
SQ SEQUENCE 381 AA; 41388 MW; 29138EBB6B4B565E CRC64;

```

Query Match 17.5%; Score 262; DB 1; Length 381;  
 Best Local Similarity 29.8%; Pred. No. 2e-14;  
 Matches 71; Conservative 42; Mismatches 87; Indels 38; Gaps 12;

```

OY 1 EDCA---ELPRRNTTEITGSMOQY-PECTQALYKCRPYR---SLGNVIMCRKGEW 53
DB 92 EFCNRSCEVPRLNRSASLKOPYITQNYFPVGTVEYECRPREPSLSPKLTCLQMLKW 151
OY 54 VALNLRKCRKPGHGDPTFGFTLTGNGVFEYGAAYVTCRNEGQQLGEINRYREDCT 113
DB 152 --STAVECKKKKSCNPEINNGQIDVGGILF--GATISSCNMTGYKLFESTS--SCLLI 206
OY 114 DG---WTNDIPICEVVKCLPVTAPENCKIVSSAMEPDR--YHFGQAVFVCSNGYKIEG 168
DB 207 SSSVQWMDPLPECREIYC-----PAPQIDNGIIGERDHYGROSVTYACNKGFTMIG 261

```

```

OY 169 DEEMHC---SDDGFSKRPKC-----VEISCKSPDVIN-----GSPISOIKIYK 210
DB 262 EHSICTVNNDEGEWSGPPPCRSKSLTSKVPPTVOKFTVNVPTVEVSPSQKTTTK 319
RESULT 14
ID LEM3_MOUSE STANDARD; PRT: 768 AA.
AC 001102;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE P-SELECTIN PRECURSOR (GRANULE MEMBRANE PROTEIN 140) (GMP-140) (PDGEM)
DE (CD62P) (LEUKOCYTE-ENDOTHELIAL CELL ADHESION MOLECULE 3) (LECAM3).
GN SELP OR GRMP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92340571.
RA Weller A., Isenmann S., Vestweber D.;
RT "Cloning of the mouse endothelial selectins. Expression of both E-
RT and P-selectin is inducible by tumor necrosis factor alpha.";
RL J. Biol. Chem. 267:15176-15183(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92345617.
RA Sanders W.E. Jr., Wilson R.W., Ballantyne C.M., Beaudet A.L.;
RT "Molecular cloning and analysis of in vivo expression of murine P-
RT selectin.";
RL Blood 80:795-800(1992).
CC -I- FUNCTION: CA(2+)-DEPENDENT RECEPTOR FOR MYELOID CELLS THAT BINDS
CC TO CARBOHYDRATES ON NEUTROPHILS AND MONOCYTES. MEDIATES THE
CC INTERACTION OF ACTIVATED ENDOTHELIAL CELLS OR PLATELETS WITH
CC LEUKOCYTES. THE LIGAND RECOGNIZED IS SIALYL-LEWIS X.
CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -I- TISSUE SPECIFICITY: STORED IN THE ALPHA-GRANULES OF PLATELETS
CC AND WEIBEL-PALADE BODIES OF ENDOTHELIAL CELLS. UPON CELL
CC ACTIVATION BY AGONISTS, P-SELECTIN IS TRANSPORTED RAPIDLY TO
CC THE CELL SURFACE.
CC -I- INDUCTION: BY TUMOR NECROSIS FACTOR ALPHA.
CC -I- SIMILARITY: TO OTHER SELECTINS/LECAMS.
CC -I- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -I- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -I- SIMILARITY: CONTAINS 8 SUSHI (SCR) REPEATS; MOUSE P-LECTIN LACKS
CC THE HUMAN SUSHI-2 EQUIVALENT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
DR EMBL: M87861; AAA40008.1; -
DR EMBL: M72332; AAA37712.1; -
DR PIR: A42755; A42755.
DR HSSP: P16109; 1FSB.
DR MGD: MGI:98280; SELP.
DR INTERPRO: IPR000436; -.
DR INTERPRO: IPR000561; -.
DR INTERPRO: IPR001304; -.
DR INTERPRO: IPR002396; -.
DR PFAM: PF00008; EGF_1.
DR PFAM: PF00059; lectin_c; 1.
DR PFAM: PF00084; sushi; 8.
DR PRINTS: PR00343; SELECTIN.
DR PROSITE: PS00022; EGF_1; 1.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE: PS00615; C_TYPE_LLECTIN_2; 1.
DR PROSITE: PS50041; C_TYPE_LLECTIN_2; 1.

```

Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;  
 KW Selectin; Signal; Sushi; Repeat; Lipoprotein; Palmitate.  
 FT CHAIN 1 41  
 FT SIGNAL 42 768  
 FT DOMAIN 42 709  
 FT TRANSMEM 710 733  
 FT DOMAIN 734 768  
 FT DOMAIN 158 195  
 FT DOMAIN 159 199  
 FT DOMAIN 199 700  
 FT REPEAT 199 700  
 FT REPEAT 258 282  
 FT REPEAT 261 330  
 FT REPEAT 323 382  
 FT REPEAT 385 444  
 FT REPEAT 447 506  
 FT REPEAT 509 568  
 FT REPEAT 579 638  
 FT REPEAT 641 700  
 FT DISULFID 60 158  
 FT DISULFID 131 150  
 FT DISULFID 163 174  
 FT DISULFID 168 183  
 FT DISULFID 185 194  
 FT DISULFID 200 244  
 FT DISULFID 230 257  
 FT DISULFID 262 306  
 FT DISULFID 292 319  
 FT DISULFID 324 368  
 FT DISULFID 354 381  
 FT DISULFID 386 430  
 FT DISULFID 416 442  
 FT DISULFID 448 493  
 FT DISULFID 478 505  
 FT DISULFID 510 554  
 FT DISULFID 540 567  
 FT DISULFID 580 624  
 FT DISULFID 610 637  
 FT DISULFID 642 686  
 FT DISULFID 672 699  
 FT CARBOHYD 398 398  
 FT CARBOHYD 603 603  
 FT CARBOHYD 654 654  
 FT CARBOHYD 661 661  
 FT CARBOHYD 679 679  
 FT LIPID 745 745  
 FT SITE 756 759  
 FT CONFLICT 724 724  
 FT SEQUENCE 768 AA: 83098 MW: E5173074D2F6E68 CRC64;

Query Match 17.4%; Score 260.5; DB 1: Length 768;  
 Best Local Similarity 26.0%; Pred. No. 5.9e-14;  
 Matches 76; Conservative 47; Mismatches 106; Indels 63; Gaps 17;  
 QY 1 EDCHLEPPRRNTTEITGWSQD-----TYPEGTQAIYKCRPG--YRSLGNVIMCR 49  
 DB 129 QDCVEIYIKSNS--APGKWNDEPGFKRRALCYTASCQDMCSNGECIETIGSYTCGY 186  
 QY 50 KG-----EWALNPLRKCK-----RPGHP-GDPTPTFTLTGNGVFEYKAVY 94  
 DB 187 PGFPGPEEYV-----KCGGVNIPQHVLMCSHPLG-----FSFNSQCTF 228  
 QY 95 TCNNGYOL--LGEINRCDTDG--WTNDIPICEVVKCLPYTAPENGRKIVSSAMEPDRHY 151  
 DB 229 SCAGYEIDGEGEL---QCLASGIWNTNPPKCAVQCGSLAPPHGNT--ACMHPIAFA 283  
 QY 152 FGCAVVRVNCNGYKIEDGEHMGSDGFGWSKEXKCVETISCKSPDV--INGG---PIQK 206  
 DB 284 YDSSCKECCPGYRARSNTLHCTGSGQWSEPLPTCEAIACEPEIPIHSGMDCVPTGT 343  
 QY 207 IYKNERFQYKCMNGYERSGDAVCTESG--W-RPLPSCEKSCDNLYIRN 256  
 DB 344 FGY--NSSCFILCAEGFVLKNGDAIQCADSGQWTAAPAFCEALDCCPEPPVS 393

RESULT 15  
 LEM2\_MOUSE  
 ID LEM2\_MOUSE STANDARD; PRT; 612 AA.  
 AC 000690;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE E-SELECTIN PRECURSOR (ENDOTHELIAL LEUKOCYTE ADHESION MOLECULE 1)  
 DE (ELAM-1) (LEUKOCYTE-ENDOTHELIAL CELL ADHESION MOLECULE 2) (LECAM2)  
 DE (CD62E).  
 GN SELE OR ELAM-1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92283265.  
 RA Becker-Andre M., van Huijsduijn R.H., Losberger C., Whelan J.,  
 RA Delamarier J.F.;  
 RT "Murine endothelial leukocyte-adhesion molecule 1 is a close  
 RT structural and functional homologue of the human protein.";  
 RL Eur. J. Biochem. 206:401-411(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92340571.  
 RA Weller A., Isenmann S., Vestweber D.;  
 RT "Cloning of the mouse endothelial selectins. Expression of both E-  
 RT and P-selectin is inducible by tumor necrosis factor alpha.";  
 RL J. Biol. Chem. 267:15176-15183(1992).  
 CC - FUNCTION: EXPRESSED ON CYTOKINE INDUCED ENDOTHELIAL CELLS AND  
 CC MEDIATES THEIR BINDING TO LEUKOCYTES. THE LIGAND RECOGNIZED BY  
 CC ELAM-1 IS SIALYL-LEWIS X (ALPHA(1->3)FUCOSYLATED DERIVATIVES OF  
 CC POLYACTOSAMINE THAT ARE FOUND AT THE NONREDUCING TERMINI OF  
 CC GLYCOLIPIDS)  
 CC - SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC - SIMILARITY: TO OTHER SELECTINS/LECAMs.  
 CC - SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
 CC - SIMILARITY: CONTAINS 6 SUSHI (SCR) REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; M80778; AAA37547.1; -;  
 CC EMBL; M87862; AAA37577.1; ALT\_INIT.  
 CC HSSP; P16581; IKAJ.  
 CC MGI; MGI:98278; SELE.  
 CC INTERPRO: IPR000436; -;  
 CC INTERPRO: IPR000561; -;  
 CC INTERPRO: IPR001304; -;  
 CC INTERPRO: IPR002396; -;  
 CC PFAM; PF00008; EGF\_1.  
 CC PFAM; PF00059; lectin\_c; 1.  
 CC PFAM; PF00084; sushi; 6.  
 CC PRINTS; PR00343; SELECTIN.  
 CC PRINTS; PS00022; BGF\_1; 1.  
 CC PROSITE; PS01186; BGF\_2; 1.  
 CC PROSITE; PS00615; C-TYPE\_LLECTIN\_1; 1.  
 CC PROSITE; PS00641; C-TYPE\_LLECTIN\_2; 1.  
 CC Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;  
 KW Selectin; Signal; Sushi; Repeat.  
 FT SIGNAL 1 21  
 FT CHAIN 22 612  
 FT DOMAIN 22 557  
 FT TRANSMEM 558 579  
 FT DOMAIN 580 612  
 E-SELECTIN.  
 EXTRACELLULAR (POTENTIAL).  
 POTENTIAL.  
 CYTOPLASMIC (POTENTIAL).

```
FT DOMAIN 38 138 C-TYPE LECTIN (SHORT FORM).
FT DOMAIN 139 175 EGF-LIKE.
FT DOMAIN 179 549 6 X SUSHI (SCR) REPEATS.
FT REPEAT 179 239 SUSHI 1.
FT REPEAT 242 301 SUSHI 2.
FT REPEAT 304 364 SUSHI 3.
FT REPEAT 367 427 SUSHI 4.
FT REPEAT 430 490 SUSHI 5.
FT REPEAT 493 549 SUSHI 6.
FT DISULFID 40 138 BY SIMILARITY.
FT DISULFID 111 130 BY SIMILARITY.
FT DISULFID 143 154 BY SIMILARITY.
FT DISULFID 148 163 BY SIMILARITY.
FT DISULFID 165 174 BY SIMILARITY.
FT DISULFID 180 225 BY SIMILARITY.
FT DISULFID 210 238 BY SIMILARITY.
FT DISULFID 243 287 BY SIMILARITY.
FT DISULFID 273 300 BY SIMILARITY.
FT DISULFID 305 350 BY SIMILARITY.
FT DISULFID 336 363 BY SIMILARITY.
FT DISULFID 368 413 BY SIMILARITY.
FT DISULFID 399 426 BY SIMILARITY.
FT DISULFID 431 476 BY SIMILARITY.
FT DISULFID 462 489 BY SIMILARITY.
FT DISULFID 494 535 BY SIMILARITY.
FT DISULFID 521 548 BY SIMILARITY.
FT CARBOHYD 25 25 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 145 145 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 192 192 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 203 203 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 266 266 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 391 391 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 504 504 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE 612 AA; 66749 MW; 86F05713F0EC2C3D CRC64;
```

## Query Match

17.4%; Score 260; DB 1; Length 612;

Best Local Similarity 27.9%; Pred. No. 5e-14; Matches 69; Conservative 36; Mismatches 120; Indels 22; Gaps 12;

```
QY 23 TYPEGTAIVKCRPGYRSLGNVIMWC-RKGEVALNPLKCRKRCRCPGHPGDTPTPGFTL 81
   :|| : : ||| : : : || : : : : : : : : : : : : : : : : : : : : :
DB 262 SYPMNTCTCTFDCVBEGRVGAQNLCCTSSGIWD--NETPSCAVTCDALIPQPNQFVSCS 319
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 82 GGNVEFYGVKAV--YTCNEGTYQLGEINIRCDPTDG-WTNDIPICEVVKCLPYTAPENG- 137
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 320 HSTAGELAFKSSCNFTCQSTFLQGPQV-ECASQGGWTFPIPVCKAVQCEALSAPOGN 378
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 138 -KIYSSAMEPDPREYHFGQAVRVCNSGKIKIGDEBMHCSDDGFWSKEKPKCVELISCKS-P 195
   | : || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 379 MKCLPFSAGP---FQNGSSCEFSCEEGFELKSRRLQCGPRGEWDSKRPKPTCSAVKCDVP 435
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 196 DYING-----SPISOKIITYKENRFOYKCNMGYEXSERGDAVCTESG-W-RPLPSCSEKS 248
   || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 436 RPQNCVMCAHAHTTEFTFYKSSCAFO--CNEGFSLHGSALQLECTISQGNKTOEVPSCQVYQ 493
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 249 CDNPYIP 255
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 494 CPSLDVP 500
```

Search completed: November 21, 2000, 16:49:21  
Job time: 324 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2000, 16:48:48 ; Search time 70.54 Seconds  
(without alignments)  
350.782 Million cell updates/sec

Title: US-09-316-163-10  
Perfect score: 1497  
Sequence: 1 EDCNELPPRRMTETITGSMSS.....EKSCDNPYIPNGDYSPLAIK 265

Scoring table:  
BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 297973 seqs, 93374136 residues

Total number of hits satisfying chosen parameters: 297973

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP invertebrate:\*  
6: SP mammal:\*  
7: SP mhc:\*  
8: SP organelle:\*  
9: SP phage:\*  
10: SP plant:\*  
11: SP rodent:\*  
12: SP virus:\*  
13: SP vertebrate:\*  
14: SP unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1497	100.0	449	4	014570
2	699	46.7	669	6	028085
3	440	29.4	1053	13	091275
4	332.5	22.2	360	12	09YT08
5	322.5	21.5	550	12	P88903
6	322.5	21.5	550	12	040912
7	320	21.4	645	12	09WRU2
8	310	20.7	259	12	P87616
9	308.5	20.6	522	6	028769
10	301	20.1	679	11	099254
11	298.5	19.9	559	4	090QV2
12	298.5	19.9	2039	4	016745
13	298.5	19.9	2489	4	016744
14	293	19.6	363	6	002839
15	290.5	19.4	315	6	028770
16	290.5	19.4	661	6	029531
17	289.5	19.3	497	6	029530
18	289.5	19.3	559	11	063612
19	289.5	19.3	559	11	063135

20	287	19.2	1911	6	029528	029528 papio hamad
21	284.5	19.0	263	12	089859	089859 variola vir
22	284.5	19.0	263	12	007033	007033 variola vir
23	283.5	18.9	263	12	089076	089076 variola vir
24	280.5	18.7	263	12	089061	089061 variola vir
25	279	18.6	1124	5	09VYR4	09VYR4 drosophila
26	277.5	18.5	560	5	022328	022328 caenorhabdi
27	276	18.4	522	6	028797	028797 pan troglod
28	274.5	18.3	349	4	015429	015429 homo sapien
29	272	18.2	285	6	019126	019126 macaca fasc
30	272	18.2	285	6	019127	019127 macaca mula
31	271	18.1	417	11	035520	035520 rattus norv
32	270.5	18.1	369	6	P79138	P79138 cercopithec
33	270.5	18.1	483	11	064735	064735 mus musculu
34	270.5	18.1	1652	5	09VIT9	09VIT9 drosophila
35	268.5	17.9	974	5	P91658	P91658 drosophila
36	267	17.8	285	6	019121	019121 papio hamad
37	266	17.8	482	6	028982	028982 sus scrofa
38	262	17.5	347	4	P78361	P78361 homo sapien
39	261	17.4	533	11	008569	008569 cavia porce
40	257	17.2	579	11	060736	060736 mus musculu
41	256.5	17.1	354	5	090611	090611 drosophila
42	256	17.1	657	4	014006	014006 homo sapien
43	255.5	17.1	1032	4	013866	013866 homo sapien
44	255	17.0	452	11	061407	061407 mus musculu
45	254.5	17.0	1045	6	046545	046545 ovis aries

## ALIGNMENTS

RESULT 1  
ID Q14570 PRELIMINARY; PRT: 449 AA.  
AC Q14570; P78435;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
DE COMPLEMENT FACTOR H PRECURSOR.  
GN HF OR CFH.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ripoche J., Day A.J., Harris T.J.R., Sim R.B.;  
RT "The complete amino acid sequence of human complement factor H.";  
RL Biochem. J. 249:593-602(1988).  
RN [2]  
RP SEQUENCE OF 226-449 FROM N.A.  
RX MEDLINE: 86169701.  
RA Kristensen T., Wetzel R.A., Tack B.F.;  
RT "Structural analysis of human complement protein H: homology with Cab  
binding protein, beta 2-glycoprotein I, and the Ba fragment of B2.";  
RL J. Immunol. 136:3407-3411(1986).  
RN [3]  
RP SEQUENCE OF 1-19 FROM N.A.  
RA Vik D.P., Williams S.A.;  
RT Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE OF 1-9 FROM N.A.  
RA Dominguez O.;  
RL Thesis (1993), Immunologia, Hospital Trias I Pujol, Spain.  
DR EMBL: X01523; CAAG0403.1; -;  
DR EMBL: M12383; AAAS2013.1; -;  
DR EMBL: U56979; AAB01987.1; -;  
DR EMBL: Z29665; CAA82763.1; -;  
DR HSSP: P10998; IYVC.  
DR INTERPRO: IPR000436; -;  
DR PRAM: PF00084; sushi; 7.  
KW SIGNAL.  
FT SIGNAL 1 18 POTENTIAL.





```

RN      (1)
RP      SEQUENCE FROM N.A.
RX      MEDLINE; 99174001.
RA      Seattles R.P., Berguian E.P., Axtheim M.K., Wong S.W.;
RT      *Sequence and genomic analysis of a Rhesus macaque rhadinovirus with
RT      similarity to Kaposi's sarcoma-associated herpesvirus/human
RL      herpesvirus 8.*;
DR      J. Virol. 73:3040-3053(1999).
EMBL: AF083501, AAD21332.1; -.
DR      INTERPRO: IPR000436; -.
PFAM: PF00084; sush1; 8.
SQ      SEQUENCE 645 AA; 71526 MW; 93D8DE35ABF61EB2 CRC64;

Query Match
Best Local Similarity 21.4%; Score 320; DB 12; Length 645;
Matches 82; Conservative 40; Mismatches 132; Indels 30; Gaps 12;

QY      1 EDCNELPPRRNTEILIGSMDSQDTPEGTQAIYKCRPGYRSLGNVIMV--CRKGEWALNP 58
DB      21 ENCK--PPHFTERYKVSNTKEKDIYSGETAELICRPGYVINTKIITTECLONGTW--STP 76
QY      59 LRKQCRPGCHPDPTPEFTTLTG--NVEYGVKAVYTCNEGYYQLGEINREC---DTD 114
DB      77 NPECDRRKCPPTADLNGAVHIGGDNALKFGSNISECEGVDLIGS--NVRFCILQDTE 135
QY      115 --GNTMDIPCEVVKCLPYTAPENCKIVSSAMEPDRE--TIFGQAVRVCNSGYKIGDEE 171
DB      136 NNNWDSNEPVCETQCKIKPPAVEHGYL----PNODYVNYGDAITFKCSLSYTLVGSTT 190
QY      172 MHCSDGFEKPKCEVSEISKSPDVING--SPISQKTIKENRERFGYCKMGYESEKGD 230
DB      191 LVCTSNKKMNSNPFICMLVCESPQIDNGYIDGLSRNRNHGSIIVKCSGDIYIVGPET 250
QY      231 AVCTESGW-RPLPSC-----EEKSCDNPIYFNGDYSPLEI 264
DB      251 LVCTNTTWPPLPKCVLVTNPSTPMPEETPMPEETPPDYOKINL 294

RESULT 8
P87616 PRELIMINARY; PRT; 259 AA.
AC      P87616.
DT      01-MAY-1997 (TREMBLrel. 03, Created)
DT      01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT      01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE      41KBP FRAGMENT FROM LEFT END OF GENOME.
GN      D17L OR C17L.
OS      Cowpox virus (CPV).
OC      Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC      Orthopoxvirus.
RN      [1]
RP      SEQUENCE FROM N.A.
RX      STRAIN=GRI-90.
RA      MEDLINE; 97068532.
RA      Saifonov P.F., Petrov N.A., Riazankina O.I., Totmenin A.V.,
RT      Shchelkunov S.N., Sandakhchiev L.S.;
RT      *Genes of a circle of hosts for the cowpox virus.*;
RT      Dokl. Akad. Nauk 349:829-833(1996).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      STRAIN=GRI-90.
RA      MEDLINE; 98229462.
RA      Shchelkunov S.N., Saifonov P.F., Totmenin A.V., Petrov N.A.,
RT      Ryzankina O.I., Gutorov V.V., Kotwal G.J.;
RT      *The genomic sequence analysis of the left and right species-specific
RT      terminal region of a cowpox virus strain reveals unique sequences and
RT      a cluster of intact ORFs for immunomodulatory and host range
RT      proteins.*;
RT      Virology 243:432-460(1998).
RL      EMBL: X94355; CAA64102.1; -.
DR      EMBL: Y11842; CAA72567.1; -.
DR      HSSP: P10998; LVCC.

```

```

DR      INTERPRO: IPR000436; -.
DR      PFAM: PF00084; sush1; 4.
SQ      SEQUENCE 259 AA; 28193 MW; 9D1AAEF6893B859A CRC64;

Query Match
Best Local Similarity 20.7%; Score 310; DB 12; Length 259;
Matches 84; Conservative 29; Mismatches 116; Indels 22; Gaps 13;

QY      3 CNELPPRRNTEILIGSMDSQDTPEGTQAIYKCRGGR--SLGNVIMCRKGEWALNPLR 60
DB      20 CCRPSRPTIMKRGKT--VDSHYNIGDTIEYLCPLPGYKQKMGPIYAKCTGTGTLFN--- 75
QY      61 KCRKRCGHPDPTPEFTTLTG--NVEYGVKAVYTCNEGYYQLGEI--NRECFDQ---W 116
DB      76 QCIKRRCPSPRIDNONGLDIGS---VDFGSSITYSKSGYHLIGESKSYCELGSTGSMW 132
QY      117 TNDIPICEVVKCLPYTAPENCKIVSSAMEPDREYHFGQAVRVCNSGYKIGDEEMHCS 176
DB      133 NPEAPICEVYKCSPPSISNGR--HNQYED--FYTDGSVYTCNSGYSILGNSGLCS- 187
QY      177 DGFMSKEPKCEVSEISKSPDVINGSPISQ--KIYKENRERQYCKMNGYESEKGDVCTE 235
DB      188 GGEWS--DPEITQIVKCPHTISNGYLSGKFKRSYSTINDVDFCKYKTLGSSSSSTCSP 246
QY      236 SG-WRP-LPSC 244
DB      247 GNTWPELPPKC 257

RESULT 9
Q28769 PRELIMINARY; PRT; 522 AA.
AC      Q28769.
DT      01-NOV-1996 (TREMBLrel. 01, Created)
DT      01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT      01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE      COMPLEMENT RECEPTOR (FRAGMENT).
OS      Papio hamadryas (Hamadryas baboon).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC      Cercopitheciinae; Papio.
RN      [1]
RP      SEQUENCE FROM N.A.
RX      TISSUE=BONE MARROW;
RA      Birmingham D.J., Logar C.M., Shen X.P., Chen W.;
RA      Submitted (MAY-1996) to the EMBL/Genbank/DBJ databases.
DR      EMBL: L77977; AAA99004.1; -.
DR      HSSP: P08603; IHFI.
DR      INTERPRO: IPR000436; -.
DR      PFAM: PF00084; sush1; 7.
FT      NON_TER 1
SQ      SEQUENCE 522 AA; 56626 MW; 312FCBE03ADF19DC CRC64;

Query Match
Best Local Similarity 20.6%; Score 308.5; DB 6; Length 522;
Matches 84; Conservative 50; Mismatches 111; Indels 37; Gaps 17;

QY      3 CN---ELPPRRNTEILIGSMDSQDTPEGTQAIYKCRPGYRSLGNVIMCRKGEWALNPL 59
DB      48 CNAPQLPFRARTNLTDAS----EPFVGYLYKECLPFGHGKPSITICLKNSVWTSACD- 102
QY      60 RKQCRPGCHPDPTPEFTTLTG--NVEYGVKAVYTCNEGYYQLGEINRECDDTG---- 115
DB      103 -KCTRKSCRNPDPVNGWVHI--KDIOFGSQINSCKMGVYLIGS--SSATCIIISGNVI 158
QY      116 WTNDIPICEVVKCLPYTAPENCKIVSSAMEPDREYH--FGQAVRVCNSG-----YITE 167
DB      159 WDNETPICEIIPCGLPPTI-ANGDFISTS---REYFPGYGVYTRCNLIGSGRKKLFELV 213
QY      168 GDEEMHCS--DD--GFWMSKEPKC-VEISKSPDVINGSPIS--QKIYKENRERFGYCKM 221

```

Db	214	GESPICYCTSKDDQYIGMSGAPQCIIIPNCKMPVENVGLVSVNRSLFSLNEVEFRCOP	273
QY	222	GYEYSEKDAVCTE-SCGRP-LPSCCEKSCDNPYPINGDISP	261
Db	274	GFWKGPGRHVOCALNWKMEPLPSCSRVOCPPPEILHGEHTP	315
RESULT	10		
Q99254	PRELIMINARY;	PRT;	679 AA.
AC	Q99254;		
DT	01-NOV-1996 (TEMBLrel. 01, Created)		
DT	01-NOV-1996 (TEMBLrel. 01, Last sequence update)		
DT	01-MAY-2000 (TEMBLrel. 13, Last annotation update)		
DE	COMPLEMENT RECEPTOR TYPE 2 (CR2) (FRAGMENT).		
GN	CR2.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-BALB/C;		
RX	MEDLINE: 90229754.		
RA	Kurtz C.B., O'Toole E., Christensen S.M., Weis J.H.;		
RT	"The murine complement receptor gene family. IV. Alternative splicing		
RT	of Cr2 gene transcripts predicts two distinct gene products that share		
RL	homologous domains with both human CR2 and CR1."		
RN	J. Immunol. 144:3581-3591(1990).		
RP	[2]		
RP	SEQUENCE OF 21-367 FROM N.A.		
RX	MEDLINE: 95105691.		
RA	Kim Y.U., Kinoshita T., Molina H., Hourcade D., Seya T., Wagner L.M.,		
RA	Holers V.M.;		
RT	"Mouse complement regulatory protein Cr1/p65 uses the specific		
RT	mechanisms of both human decay-accelerating factor and membrane		
RT	cofactor protein."		
RL	J. Exp. Med. 181:151-159(1995).		
DR	EMBL: U17126; AAA78271.1; JOINED.		
DR	EMBL: U17123; AAA78271.1; JOINED.		
DR	EMBL: U17124; AAA78271.1; JOINED.		
DR	EMBL: U17125; AAA78271.1; JOINED.		
DR	EMBL: U17126; AAA78271.1; JOINED.		
DR	EMBL: U17127; AAA78271.1; JOINED.		
DR	EMBL: M36470; AAA37449.1; -.		
DR	HSSP: P10998; IVC.		
DR	MGSP: MGI:88489; Cr2.		
DR	INTERPRO: IPR000436; -.		
DR	PFAM: PF00084; sushi; 10.		
FT	NON_TER 679		
FT	SEQUENCE 679 AA; 74916 MW; 52PC00FDCED20CDC CXC64;		
Query Match	20.1%;	Score 301;	DB 11; Length 679;
Best Local Similarity	26.0%;	Pred. No. 7.6e-20;	
Matches	82;	Conservative 51;	Mismatches 94; Indels 88; Gaps 16;
QY	20	SDQT-YPEGTQAIVKCRPGYRSLGNVIMYCRKGEVVALNPLRCKCRPCGHPGDPFGTFE	78
Db	37	SDKSEFAIGTWEYKCRPGYFRKSPFLINCLFETSKW--SDAQFCRKRKPCMPNPOEPLHGSV	94
QY	79	TLTGNGVEFYKAVYITTCNEGYQLLGEINRRC-----PDTGNTNDIPICEVYKCLPYTAP	134
Db	95	HINTG--IEFGSTIYSNCGYRLIGD--SSATCIYSDNTVMMDNMDPLESIPCESPPAI	151
QY	135	ENGKIVSSAMEDDREKHFQGAARFVPCNSG-----YKIGDEEMHCSD-----GFWSKER	184
Db	152	SNGDYSSSRD---SFFYGMVYTYCHTKRNEKFLDLVGEKSIYCTSDNDNOYGINSP	208
QY	185	PKCV-EISCKSPDVING-----	204
Db	209	PCQIPRVKCPMEIEINGLVESGFKHSFPLNDTVIFKCKSGFTMKGSRIAMQPNKSWSP	268
QY	205	-----OKIIT-----KENERP-----QYCKNMGYEYSEKDAVCTESG-W-RDL	241

Db	269	LPCTCMGLCPONTILHGDYNNKDEFFSVGOKVSTYCNBNGYTLTGINTLIVECTSLGTWSMTV	328
QY	242	PSCEBSCDNPYTPN 256	
Db	329	PTCEVKSCD--AIPN 341	
RESULT	11		
ID	Q90QV2	PRELIMINARY; PRT; 559 AA.	
AC	Q90QV2;		
DT	01-MAY-2000 (TREMBLrel. 13, created)		
DT	01-MAY-2000 (TREMBLrel. 13, last sequence update)		
DT	01-JUN-2000 (TREMBLrel. 14, last annotation update)		
OS	HUMAN CRI MRNA FOR C3B/C4B RECEPTOR SECRETED FORM PRECURSOR (FRAGMENT).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEIDLIN; 89010527.		
RA	Hourcade D., Mesner D.R., Atkinson J.P., Holers V.M.;		
RT	"Identification of an alternative polyadenylation site in the human		
RT	C3b/C4b receptor (complement receptor type 1) transcriptional unit and		
RL	prediction of a secreted form of complement receptor type 1.";		
DR	J. Exp. Med. 168:1255-1270(1988).		
DR	EMBL; X14362; CAA32541.1; -.		
DR	INTERPRO; IPR000436; -.		
DR	INTERPRO; IPR000834; -.		
DR	PFAM; PF00084; sushi; 8.		
DR	PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_1.		
KW	Signal.		
FT	NON_TER	1	
FT	SIGNAL	<1 16	POTENTIAL.
FT	CHAIN	17 559	POTENTIAL.
SEQUENCE	559 AA; 61424 MW; DBFFE965CA179D75 CRC64;		
Query Match	19.9%;	Score 298.5;	DB 4; Length 559;
Best Local Similarity	27.0%;	Pred. No. 1e-19;	
Matches	86; Conservative	49; Mismatches	111; Indels 73; Gaps 18
QY	3	CNE---LPRRRTTELTGWSMDQYDEPGQALYKCPGRSLGNVIMVCRKGEMVALNPL 59	
Db	18	CAPEMLPFPARPTNLN---DEFEFFIGYLTETCRPGISGRPFSLTCLKNSVYTGAKD- 72	
QY	60	RKCOKRKPCHPDPTFGFTFLTGANVFEGYKAVAYCNGSYOLGELINYREC---DTDG 115	
Db	73	KCRRSKCNPNPDPPVAGMNVHVIK--IQGSGQIKYSCCTGYNRLIGS-SSATCGLISGDYI 128	
QY	116	WTNDIPICEVVKC-LFVTAPENGKIVYSAMEPDREYHFGQAVRFVCNSG-----KIEG 168	
Db	129	WNEPEPICRIPGCLPPT-TNGDFISTNRE--NFHYGSVVTYRCNPGSGGRKVELWG 184	
QY	169	DEEMHC--SDD--GFMRSKEKPC-VISCKSPDVINGSPIS-QKILYKNEROYKCNMG 222	
Db	185	EPSIYTSNDVDGIVMSGAPQCIIPNCKTPPVENGILVSDNRSLFSLNEVVEFRCQPG 244	
QY	223	YEKSEKGAVDCTE-SGMRP-LPSCDE-----KSCDNP 252	
Db	245	FVWKGRPRVKCAQALNNEPELIPGCSKVCQPPDYLIAERTQRDKDNFSPQGVFYSCERG 304	
QY	253	y-----IPNGDYS 261	
Db	305	YDLRGAASMRCTPGDWSP 323	
RESULT	12		
ID	Q16745	PRELIMINARY; PRT; 2039 AA.	
AC	Q16745;		
DT	01-NOV-1996 (TREMBLrel. 01, created)		

```
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE COMPLEMENT RECEPTOR 1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
RN Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.
RP SEQUENCE FROM N.A.
RX MEDLINE: 94065175.
RA Vik D.P., Wong W.W.;
RT "Structure of the gene for the F allele of complement receptor type 1
RT and sequence of the coding region unique to the S allele."
J. Immunol. 151:6214-6224(1993).
[2]
RP SEQUENCE FROM N.A.
RA Vik D.P., Wong W.W.;
DT Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; L17418; AAB60694.1; -.
DR EMBL; L17390; AAB60694.1; JOINED.
DR EMBL; L17391; AAB60694.1; JOINED.
DR EMBL; L17392; AAB60694.1; JOINED.
DR EMBL; L17393; AAB60694.1; JOINED.
DR EMBL; L17394; AAB60694.1; JOINED.
DR EMBL; L17395; AAB60694.1; JOINED.
DR EMBL; L17396; AAB60694.1; JOINED.
DR EMBL; L17397; AAB60694.1; JOINED.
DR EMBL; L17398; AAB60694.1; JOINED.
DR EMBL; L17399; AAB60694.1; JOINED.
DR EMBL; L17400; AAB60694.1; JOINED.
DR EMBL; L17401; AAB60694.1; JOINED.
DR EMBL; L17402; AAB60694.1; JOINED.
DR EMBL; L17403; AAB60694.1; JOINED.
DR EMBL; L17404; AAB60694.1; JOINED.
DR EMBL; L17405; AAB60694.1; JOINED.
DR EMBL; L17406; AAB60694.1; JOINED.
DR EMBL; L17407; AAB60694.1; JOINED.
DR EMBL; L17408; AAB60694.1; JOINED.
DR EMBL; L17409; AAB60694.1; JOINED.
DR EMBL; L17410; AAB60694.1; JOINED.
DR EMBL; L17411; AAB60694.1; JOINED.
DR EMBL; L17412; AAB60694.1; JOINED.
DR EMBL; L17413; AAB60694.1; JOINED.
DR EMBL; L17414; AAB60694.1; JOINED.
DR EMBL; L17415; AAB60694.1; JOINED.
DR EMBL; L17416; AAB60694.1; JOINED.
DR EMBL; L17417; AAB60694.1; JOINED.
DR EMBL; L17419; AAB60694.1; JOINED.
DR EMBL; L17420; AAB60694.1; JOINED.
DR EMBL; L17421; AAB60694.1; JOINED.
DR EMBL; L17422; AAB60694.1; JOINED.
DR EMBL; L17423; AAB60694.1; JOINED.
DR HSSP; P08603; HFT.
DR INTERPRO; IPR000436; -.
DR INTERPRO; IPR000834; -.
DR INTERPRO; IPR001424; -.
DR PFM; PF000084; sushi; 30.
DR PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN_1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_2.
SQ SEQUENCE 2039 AA; 223603 MW; B82FCB11C6B16635 CRC64;
```

```
Query Match 19.9%; Score 298.5; DB 4; Length 2039;
Best Local Similarity 27.0%; Pred. No. 5e-19;
Matches 86; Conservative 49; Mismatches 111; Indels 73; Gaps 18;
```

```
OY 3 CNE---LPPRRTTELITGMSQITPEGQALIKCRPGTRISGNVIMVCRKGEWALNPL 59
DB 43 CNAPEWLPAPRTNLNT---DEFEPPIGTYLMECRPGTSGRPFSTICLKNVWTGAKD- 97
OY 60 RCKCRPCGHDPDPFGFTTTLTGAVFEGVAVYTCNGEYOLLEGINREC---DTG 115
DB 98 RCRKRSCRNPPDPVNGAVHVIKQ---IQFGSOIKISCTKGYRLIGS--SATCIIISGDIYI 153
```

```
OY 116 WTNDIPICEVVK-LPYTAPENGKIVSAMPEDEYHFGAVREVCNSG-----YKIEG 168
DB 154 WDNETPICDRIPCLPPTI-TNGDFISTNRE---NFHYGSVVTYRCNNGSGGRVFEVLVG 209
OY 169 DEEMHC--SDD--GFMSKEKPKC-VEISCKSPDVINSPTS-QATIIKENRFFQYCKNMG 222
DB 210 EPSTICTSNDQGVIGMSPPAQCTIPNKCCTPPNVENGILVSDNLSFLNEVEFECPG 269
OY 223 YEYSEKGDVACTE-SGNRP-LPSCSE-----KSCDNP 252
DB 270 FVMKGPFRVQKQALNKKMEPELPSCSRVCCQPPPDVIAHERRQKDNPSPOGEVRYSCPC 329
OY 253 Y-----IPNGDYSP 261
DB 330 YDLRGAASMRCTPOGDWSP 348

RESULT 13
ID 016744 PRELIMINARY; PRT; 2489 AA.
AC 016744;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE COMPLEMENT RECEPTOR 1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
RN Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.
RP SEQUENCE FROM N.A.
RX MEDLINE: 94065175.
RA Vik D.P., Wong W.W.;
RT "Structure of the gene for the F allele of complement receptor type 1
RT and sequence of the coding region unique to the S allele."
J. Immunol. 151:6214-6224(1993).
[2]
RP SEQUENCE FROM N.A.
RA Vik D.P., Wong W.W.;
DT Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; L17418; AAB60695.1; -.
DR EMBL; L17390; AAB60695.1; JOINED.
DR EMBL; L17391; AAB60695.1; JOINED.
DR EMBL; L17392; AAB60695.1; JOINED.
DR EMBL; L17393; AAB60695.1; JOINED.
DR EMBL; L17394; AAB60695.1; JOINED.
DR EMBL; L17395; AAB60695.1; JOINED.
DR EMBL; L17396; AAB60695.1; JOINED.
DR EMBL; L17397; AAB60695.1; JOINED.
DR EMBL; L17398; AAB60695.1; JOINED.
DR EMBL; L17399; AAB60695.1; JOINED.
DR EMBL; L17400; AAB60695.1; JOINED.
DR EMBL; L17401; AAB60695.1; JOINED.
DR EMBL; L17402; AAB60695.1; JOINED.
DR EMBL; L17403; AAB60695.1; JOINED.
DR EMBL; L17405; AAB60695.1; JOINED.
DR EMBL; L17406; AAB60695.1; JOINED.
DR EMBL; L17407; AAB60695.1; JOINED.
DR EMBL; L17408; AAB60695.1; JOINED.
DR EMBL; L17409; AAB60695.1; JOINED.
DR EMBL; L17410; AAB60695.1; JOINED.
DR EMBL; L17411; AAB60695.1; JOINED.
DR EMBL; L17412; AAB60695.1; JOINED.
DR EMBL; L17413; AAB60695.1; JOINED.
DR EMBL; L17414; AAB60695.1; JOINED.
DR EMBL; L17415; AAB60695.1; JOINED.
DR EMBL; L17416; AAB60695.1; JOINED.
DR EMBL; L17417; AAB60695.1; JOINED.
DR EMBL; L17419; AAB60695.1; JOINED.
DR EMBL; L17420; AAB60695.1; JOINED.
DR EMBL; L17421; AAB60695.1; JOINED.
DR EMBL; L17422; AAB60695.1; JOINED.
DR EMBL; L17423; AAB60695.1; JOINED.
DR EMBL; L17424; AAB60695.1; JOINED.
```



QY 255 -----PNGDYSP 261  
| | | |  
Db 287 PSYDLRGASLRCTPOGDWSP 307

Search completed: November 21, 2000, 16:48:51  
Job time: 329 sec

PT by modulating or detecting tumour associated human complement factors

PT H related antigen, or nucleic acid encoding it  
XX  
XX Example 6B; Fig 6B; 104pp; English.  
CC This partial protein sequence represents a region of the human  
CC tumour-associated complement factor H (CFH). This sequence is used  
CC in the identification of complement factor H related proteins and  
CC antigens isolated from clone pRAB9FH410 (see W39155). The detection of  
CC such proteins and a CFH antigens can be used in screening or for the  
CC treatment of renal or urogenital cancer, e.g. bladder, cervical or  
CC prostate cancer. Agents that may modulate this antigen could be used in  
CC the manufacture of a medicament for the treatment of a tumour cell.  
SQ Sequence 240 AA;  
  
Query Match 62.0%; Score 721; DB 18; Length 240;  
Best Local Similarity 100.0%; Pred. No. 1.2e-59;  
Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 78 FTLTGNGVEYGYKAYVTCNEGYOLLGEINVRCDTGDGTNDIPICGVKCLPTAPENG 137  
DB 1 fcltgnvteygykavvtcmegyllgelnytecdtdgwtndipicevkcclptapeng 60  
  
QY 138 KIVSSAMEPDRHYHFGQAVFVNCNSGYKICDEMHCSDDGFWSKRPKCVETISCKSPDV 197  
DB 61 kivsamedprehyhfgqavrfvncnsgykiedemhcsddgfwskepkcveiscskspdv 120  
  
QY 198 INGSPISSOKI 207  
DB 121 ingspslsqk1 130  
  
RESULT 2  
W39155  
ID W39155 standard; Protein: 216 AA.  
XX W39155;  
AC W39155;  
XX  
XX 27-APR-1998 (first entry)  
DT  
XX  
XX Clone pRAB9FH410 CFH related protein fragment.  
DE  
XX  
XX Complement factor H; tumour associated antigen; renal cancer;  
KW urogenital cancer; medicament; modulator.  
OS Synthetic.  
XX  
XX W09738136-A1.  
PN  
XX  
XX 16-OCT-1997.  
PD  
XX  
XX 09-APR-1997; 97WO-US05710.  
PF  
XX  
XX 06-MAR-1997; 97US-0812481.  
PR  
XX 09-APR-1996; 96US-0015083.  
PR  
XX 09-APR-1996; 96US-0630048.  
PR  
XX 06-MAR-1997; 97US-0038614.  
PR  
XX  
XX (BARD-) BARD DIAGNOSTIC SCI INC.  
PA  
XX  
XX Entfield DL, Hass GM, Kinders RJ;  
PI  
XX  
XX WPI; 1997-512742/47.  
DR  
XX  
XX N-PSDB; V02791.  
DR  
XX  
XX Treating or screening for cancer, e.g. renal or urogenital cancer -  
PT by modulating or detecting tumour associated human Complement Factor  
PT H related antigen, or nucleic acid encoding it  
XX  
XX Example 6B; Fig 6B; 104pp; English.  
XX  
XX This partial protein is found in clone pRAB9FH410 and represents a

CC complement factor H related protein with homology to a region of the  
CC human tumour-associated complement factor H (CFH). The detection of this  
CC protein and a CFH antigen can be used in screening or for the treatment  
CC of renal or urogenital cancer, e.g. bladder, cervical or prostate cancer.  
CC Agents that may modulate this antigen could be used in the manufacture of  
CC a medicament for the treatment of a tumour cell.  
SQ Sequence 216 AA;  
  
Query Match 53.9%; Score 627; DB 18; Length 216;  
Best Local Similarity 99.1%; Pred. No. 6e-51;  
Matches 112; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 95 TCNEGYOLLGEINVRCDTGDGTNDIPICGVKCLPTAPENGKIVSSAMEPDRHYHFGQ 154  
DB 1 tcnegyollgelnytecdtdgwtndipicevkcclptapengkivsamedprehyhfg 60  
  
QY 155 AVFVNCNSGYKICDEMHCSDDGFWSKRPKCVETISCKSPDVINGSPISOKI 207  
DB 61 avrfvncnsgykiedemhcsddgfwskepkcveiscskspdvingspslsqk1 113  
  
RESULT 3  
W06881  
ID W06881 standard; Protein: 299 AA.  
XX W06881;  
AC W06881;  
XX  
XX 18-MAR-1997 (first entry)  
DT  
XX  
XX Decay accelerating factor.  
DE  
XX  
XX Complement inhibitor; membrane co-factor protein; MCP;  
KW decay accelerating factor; DAF; chimeric protein; glycosaminoglycan;  
KW heparin; cell lysis; sepsis; adult respiratory distress syndrome;  
KW reperfusion injury; cell damage.  
XX  
XX Homo sapiens.  
OS  
XX  
XX W09634965-A2.  
PN  
XX  
XX 07-NOV-1996.  
PD  
XX  
XX 03-MAY-1996; 96WO-US06301.  
PF  
XX  
XX 05-MAY-1995; 95US-0435149.  
PR  
XX  
XX (CHIR ) CHIRON CORP.  
PA  
XX  
XX Creasey AA, Innis MA, Zaror I;  
PI  
XX  
XX WPI; 1996-506167/50.  
DR  
XX  
XX Chimeric proteins for inhibiting complement-mediated cell lysis -  
PT comprise membrane co-factor protein and decay accelerating factor  
PT peptide sequences  
XX  
XX  
XX Disclosure: Page 25; 33pp; English.  
XX  
XX A portion (W06881) of the complement-inhibitor, decay accelerating  
CC factor (DAF), is used in novel chimeric proteins of formula  
CC A-R1-B-R2-C, where A and C are peptides (W06875-79, W06883-90) able  
CC to bind glycosaminoglycans (esp. heparin) present on cell surfaces,  
CC R1 is a portion of DAF or membrane co-factor protein (MCP, see also  
CC W06880), R2 is DAF when R1 is MCP or MCP when R1 is DAF, and B is a  
CC peptide that may have complement inhibitor activity. The chimeric  
CC proteins (see also W06882) are directed to cell surfaces where they  
CC inhibit complement-mediated cell lysis. They are used to treat and  
CC prevent disease states in which complement plays a role, e.g.  
CC sepsis, adult respiratory distress syndrome, reperfusion injury and  
CC tissue damage.  
XX



[illegible]

Dd	92	eefnicsecevrptlnasalskpyltqnyfvygtvvecyecrrprirrepsaklcltqlnlkw
Oy	54	VALNPLRKQCKRRCCGHPGDTPEFGTTLVLGGNVFEYGVAAVTCDEGYOLLGEINWRECDT
Dd	152	--stavefcckksccppgeirlngqidvpqgllf--gatlsfscntlykllfgsts-sfccl
Oy	114	DG----WTNDIPICEVVKCLPTAPENCKIVTSAMEPPE- YHGGGAVPFNCNGYKIEG
Dd	207	sagsvwssplplectrelc-----papqdldglqgerdlhygrsvtyacnkgitumig
Oy	169	DELAMHC---SDGCFMSKEPKC-----VEISCKSPDVIN-----GSPISK
Dd	262	ebsiyctvnndegewsgpppecrqrsktskypvtvgkxpltvnpvtleavspstsqk
RESULT	7	
ID	R66683	standard; Protein; 381 AA.
XX	R66683;	
DT	23-JUL-1995	(first entry)
XX	DE	Decay accelerating factor.
XX	KX	Decay accelerating factor; DAF; mDAF; fusion protein; liposome;
XX	KW	cell targeting; glycoposphatidylinositol; GPI; drug delivery.
XX	OS	Homo sapiens.
FT	Key	Location/Qualifiers
FT	Peptide	1..34
FT	Modified-site	/label= Sig-peptide
FT		364
FT		/note= "probable phosphatidylinositol
PN	US5374548-A.	derivatization site"
PD	20-DEC-1994.	
PF	02-MAY-1986;	86US-0859107.
PR	02-MAY-1986;	86US-0859107.
PR	06-AUG-1987;	87US-0083757.
PR	19-DEC-1991;	91US-0011048.
PR	12-FEB-1993;	93US-0017934.
PA	(GETH ) GENENTECH INC.	
PI	Caras IW:	
DR	WPI, 1995-035649/05.	
DR	N-PsDB: Q79863.	
Liposome(s)	for targetting particular cells contg. fusion protein	
- of glyco.phosphatidyl.inositol anchor and heterologous	targeting protein, e.g. for delivering toxins to infected or	
cancer cells.		
Disclosure; Fig. 1a-1f; 36pp; English.		
A probe (given in Q79865) based on the N-terminal sequence of human		
decay accelerating factor (DAF) was used to screen a HeLa cell		
lambda cDNA library. Isolated clones encoding membrane-bound DAF		
(mDAF) were obtained, and the full sequence of mDAF cDNA was		
determined (Q79865). The GPI signal domain of mDAF may be fused to		
a heterologous protein and targeted to cell membrane surfaces.		
Sequence 381 AA;		

Query Match      22.4% ; Score 261; DB 16; Length 381;







(e.g. kits for determining whether a

XX

WPI: 1992-375009/46





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2000, 16:43:57 ; Search time 27.94 Seconds  
(without alignments)  
236.694 Million cell updates/sec

Title: US-09-316-163-9

Perfect score: 1163  
Sequence: 1 EDCNLEPPRRNTEILGWSMS.....VEISCKSPDVINGSPISQKI 207

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 87993 seqs, 31947931 residues

Total number of hits satisfying chosen parameters: 87993

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first=45 summaries

Database: SwissProt\_39\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1163	100.0	1231	1 CFAH_HUMAN	P08603 homo sapien
2	822	70.7	1234	1 CFAH_MOUSE	P06909 mus musculu
3	261	22.4	381	1 DAF_HUMAN	P08174 homo sapien
4	255	21.9	2039	1 CRI_HUMAN	P17927 homo sapien
5	253.5	21.8	484	1 LEM2_PIG	P98110 sus scrofa
6	251	21.6	340	1 DAF_PONPY	P49457 pongo pygma
7	246.5	21.2	507	1 DAF_CAVPO	Q60401 cavia porce
8	246	21.2	263	1 VCP_VACCV	P10998 vaccinia vi
9	244	21.0	597	1 C4BP_HUMAN	P04003 homo sapien
10	238.5	20.5	377	1 MCP_HUMAN	P15529 homo sapien
11	236	20.3	360	1 CCPH_HSVSA	Q01016 herpesvirus
12	231.5	19.9	390	1 DAF1_MOUSE	Q61475 mus musculu
13	230.5	19.8	610	1 LEM2_HUMAN	P16581 homo sapien
14	228	19.6	469	1 C4BP_MOUSE	P08607 mus musculu
15	227.5	19.6	768	1 LEM3_MOUSE	Q01102 mus musculu
16	227.5	19.6	830	1 LEM3_HUMAN	P16109 homo sapien
17	226	19.4	549	1 LEM2_RAT	P98105 rattus norv
18	222.5	19.1	407	1 DAF2_MOUSE	Q61476 mus musculu
19	222	19.1	768	1 LEM3_RAT	P98106 rattus norv
20	220	18.9	611	1 LEM2_CANFA	P33730 canis fami
21	218.5	18.8	612	1 LEM2_MOUSE	Q00690 mus musculu
22	216	18.6	558	1 C4BP_RAT	P63514 rattus norv
23	214.5	18.4	551	1 LEM2_RABIT	P27113 oryctolagus
24	214	18.4	610	1 C4BP_BOVIN	Q28065 bos taurus
25	212	18.2	1025	1 LEM3_SHEEP	P98109 ovis aries
26	212	18.2	769	1 CR2_MOUSE	P19070 mus musculu
27	210.5	18.1	646	1 LEM3_BOVIN	P42201 bos taurus
28	210	18.1	1033	1 CR2_HUMAN	P20023 homo sapien
29	210	18.1	372	1 LEM1_MOUSE	P18337 bos musculu
30	205.5	17.7	465	1 LEM2_BOVIN	P98107 bos taurus
31	205	17.6	345	1 APOH_BOVIN	P17650 bos taurus
32	202	17.4	958	1 HIG_DROME	Q09101 drosophila
33	200	17.2	372	1 LEM1_RAT	P30836 rattus norv

34	199	17.1	345	1 APOH_MOUSE	Q01339 mus musculu
35	194	16.7	372	1 LEM1_HUMAN	P14151 homo sapien
36	194	16.7	372	1 LEM1_MACMU	Q95198 macaca mula
37	194	16.7	372	1 LEM1_PANTR	Q95237 pan troglod
38	194	16.7	372	1 LEM1_PAPHA	Q28766 papio hamad
39	193	16.6	372	1 LEM1_PONPY	Q95233 pongo pygma
40	191.5	16.5	661	1 F13B_HUMAN	P05160 homo sapien
41	188	16.2	258	1 C4BP_RAT	Q63515 rattus norv
42	188	16.2	1019	1 LFC_TACR	P28175 tachyples
43	187.5	16.1	345	1 APOH_HUMAN	P02749 homo sapien
44	184	15.8	1019	1 LFC_CARRO	Q26422 carlinoscor
45	182	15.6	345	1 APOH_CANFA	P33703 canis fami

## ALIGNMENTS

RESULT 1  
ID CFAH\_HUMAN STANDARD: PRT: 1231 AA.  
AC P08603; 01-AUG-1988 (rel. 08, Created)  
DT 01-JAN-1990 (rel. 13, Last sequence update)  
DI 30-MAY-2000 (rel. 39, Last annotation update)  
DE COMPLEMENT FACTOR H PRECURSOR.  
GN HFI OR HF OR CFH.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER:  
RX MEDLINE: 86134059.  
RA Ripchoe J., Day A.J., Harris T.J.R., Sim R.B.;  
RT "The complete amino acid sequence of human complement factor H.";  
RL Biochem. J. 249:593-602(1988).  
RN [2]  
RP SEQUENCE OF 53-445 FROM N.A.  
RX MEDLINE: 87054207.  
RA Schulz T.F., Schwaeble W., Stanley K.K., Weiss E., Dierich M.P.;  
RT "Human complement factor H: Isolation of cDNA clones and partial cDNA  
sequence of the 38-kDa tryptic fragment containing the binding site  
for C3b.";  
RL Eur. J. Immunol. 16:1351-1355(1986).  
RN [3]  
RP SEQUENCE OF 226-445 FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE: 86169701.  
RA Kristensen T., Wetsel R.A., Tack B.F.;  
RT "Structural analysis of human complement protein H: homology with C4b  
binding protein, beta 2-glycoprotein I, and the Ba fragment of B2.";  
RL J. Immunol. 136:3407-3411(1986).  
RN [4]  
RP SEQUENCE OF 1047-1231 FROM N.A.  
RX MEDLINE: 91201892.  
RA Estallier C., Kristensen V., Schwaeble W., Dierich M.P., Weiss E.H.;  
RT "Cloning of the 1.4-kb mRNA species of human complement factor H  
reveals a novel member of the short consensus repeat family related  
to the carboxy terminal of the classical 150-kDa molecule.";  
RL J. Immunol. 146:3190-3196(1991).  
RN [5]  
RP SEQUENCE OF 19-35.  
RX MEDLINE: 83048213.  
RA Sim R.B., Disclipio R.G.;  
RT "Purification and structural studies on the complement-system control  
protein beta 1H (Factor H).";  
RL Biochem. J. 205:285-293(1982).  
RN [6]  
RP STRUCTURE BY NMR OF 927-985 (SUSHI 16).  
RX MEDLINE: 91278097.  
RA Norman D.G., Barlow P.N., Baron M., Day A.J., Sim B., Campbell I.D.;  
RT "three-dimensional structure of a complement control protein module  
in solution.";  
RL J. Mol. Biol. 219:717-725(1991).

[7]  
RN STRUCTURE BY NMR OF 264-322 (SUSHI 5).  
RP MEDLINE; 92232649.  
RA Barlow P.N., Norman D.G., Steinkasserer A., Horne T.J., Pearce J.,  
RT "Solution structure of the fifth repeat of factor H: a second example  
of the complement control protein module.";  
RL Biochemistry 31:3626-3634(1992).  
[8]  
RN STRUCTURE BY NMR OF 866-985 (SUSHI 15 AND 16).  
RP MEDLINE; 93323119.  
RA Barlow P.N., Steinkasserer A., Norman D.G., Kieffer B., Wiles A.P.,  
RT "Solution structure of a pair of complement modules by nuclear  
magnetic resonance.";  
RL J. Mol. Biol. 232:268-284(1993).  
-1- FUNCTION: FACTOR H FUNCTIONS AS A COFACTOR IN THE INACTIVATION OF  
C3B BY FACTOR I AND ALSO INCREASES THE RATE OF DISSOCIATION OF THE  
C3BBB COMPLEX (C3 CONVERTASE) AND THE (C3B)NBB COMPLEX (C5  
CONVERTASE) IN THE ALTERNATIVE COMPLEMENT PATHWAY.  
-1- SIMILARITY: CONTAINS 20 SUSHI (SCR) REPEATS.  
-1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION  
341 ONWARD DUE TO A FRAMESHIFT.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; Y00716; CAA68704.1; -;  
DR EMBL; X04697; CAA61739.1; ALT\_FRAME.  
DR EMBL; M65294; AAA35948.1; -;  
DR PIR; S00254; NBUH.  
DR PIR; S03013; S03013.  
DR PIR; S00254; S00254.  
DR PDB; IHCC; 15-APR-92.  
DR PDB; IHP1; 15-JUL-93.  
DR MIM; 134370; -;  
DR INTERPRO: IPR000436; -;  
DR PFAM: PF00084; sushi. 20.  
KW Complement alternate pathway; Plasma; Glycoprotein; Repeat; Sushi;  
Signal; 3d-structure; Polymorphism.  
FT SIGNAL 1 18  
FT CHAIN 19 1231  
FT DOMAIN 20 1229  
FT REPEAT 20 81  
FT REPEAT 84 142  
FT REPEAT 145 206  
FT REPEAT 209 263  
FT REPEAT 266 321  
FT REPEAT 324 386  
FT REPEAT 388 443  
FT REPEAT 447 506  
FT REPEAT 508 565  
FT REPEAT 568 624  
FT REPEAT 629 685  
FT REPEAT 690 745  
FT REPEAT 752 804  
FT REPEAT 810 865  
FT REPEAT 869 927  
FT REPEAT 930 985  
FT REPEAT 988 1044  
FT REPEAT 1047 1103  
FT REPEAT 1108 1164  
FT REPEAT 1166 1229  
FT DISULFID 21 66  
FT DISULFID 52 80  
FT DISULFID 85 129  
FT DISULFID 114 141  
COMPLEMENT FACTOR H.  
20 X SUSHI (SCR) REPEATS.  
SUSHI 1.  
SUSHI 2.  
SUSHI 3.  
SUSHI 4.  
SUSHI 5.  
SUSHI 6.  
SUSHI 7.  
SUSHI 8.  
SUSHI 9.  
SUSHI 10.  
SUSHI 11.  
SUSHI 12.  
SUSHI 13.  
SUSHI 14.  
SUSHI 15.  
SUSHI 16.  
SUSHI 17.  
SUSHI 18.  
SUSHI 19.  
SUSHI 20.  
BY SIMILARITY.  
BY SIMILARITY.  
BY SIMILARITY.  
BY SIMILARITY.  
BY SIMILARITY.

FT	DISULFID	146	192	BY SIMILARITY.
FT	DISULFID	178	205	BY SIMILARITY.
FT	DISULFID	210	251	BY SIMILARITY.
FT	DISULFID	237	262	BY SIMILARITY.
FT	DISULFID	267	309	BY SIMILARITY.
FT	DISULFID	294	320	BY SIMILARITY.
FT	DISULFID	325	374	BY SIMILARITY.
FT	DISULFID	357	385	BY SIMILARITY.
FT	DISULFID	389	431	BY SIMILARITY.
FT	DISULFID	416	442	BY SIMILARITY.
FT	DISULFID	448	494	BY SIMILARITY.
FT	DISULFID	477	505	BY SIMILARITY.
FT	DISULFID	509	553	BY SIMILARITY.
FT	DISULFID	536	564	BY SIMILARITY.
FT	DISULFID	567	611	BY SIMILARITY.
FT	DISULFID	597	623	BY SIMILARITY.
FT	DISULFID	630	673	BY SIMILARITY.
FT	DISULFID	659	684	BY SIMILARITY.
FT	DISULFID	691	733	BY SIMILARITY.
FT	DISULFID	719	744	BY SIMILARITY.
FT	DISULFID	753	792	BY SIMILARITY.
FT	DISULFID	781	803	BY SIMILARITY.
FT	DISULFID	811	853	BY SIMILARITY.
FT	DISULFID	839	864	BY SIMILARITY.
FT	DISULFID	870	915	BY SIMILARITY.
FT	DISULFID	901	926	BY SIMILARITY.
FT	DISULFID	931	973	BY SIMILARITY.
FT	DISULFID	959	984	BY SIMILARITY.
FT	DISULFID	989	1032	BY SIMILARITY.
FT	DISULFID	1018	1043	BY SIMILARITY.
FT	DISULFID	1048	1091	BY SIMILARITY.
FT	DISULFID	1077	1102	BY SIMILARITY.
FT	DISULFID	1109	1152	BY SIMILARITY.
FT	DISULFID	1138	1163	BY SIMILARITY.
FT	DISULFID	1167	1218	BY SIMILARITY.
FT	DISULFID	1201	1228	BY SIMILARITY.
FT	CARBOHYD	529	529	BY SIMILARITY.
FT	CARBOHYD	718	718	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	802	802	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	822	822	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	882	882	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	911	911	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1029	1029	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1095	1095	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARIANT	402	402	H -> Y.
FT	CONFLICT	21	21	/FTID-VAR_001979.
FT	CONFLICT	30	30	C -> Q (IN REF. 3).
FT	CONFLICT	34	34	T -> V (IN REF. 3).
FT	CONFLICT	53	54	T -> Q (IN REF. 3).
FT	CONFLICT	870	871	RP -> IL (IN REF. 2).
FT	STRAND	876	876	
FT	TURN	877	878	
FT	STRAND	881	881	
FT	STRAND	890	891	
FT	TURN	893	894	
FT	STRAND	896	900	
FT	TURN	903	904	
FT	STRAND	907	907	
FT	TURN	912	916	
FT	STRAND	917	918	
FT	STRAND	919	920	
FT	STRAND	926	926	
SO	SEQUENCE	1231 AA;	139125 MW;	C65EC8CF800B3FD CRC64;

Query Match 100.0%; Score 1163; DB 1; Length 1231;  
Best Local Similarity 100.0%; Pred. No. 1,2e-94;  
Matches 207; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 EDCNLEPPRRNTEITLIGSWSDQTYPEGTQAIYKCRPGYRSLGNVIMVCRKGEMVALNPLR 60  
DB 19 EDCNLEPPRRNTEITLIGSWSDQTYPEGTQAIYKCRPGYRSLGNVIMVCRKGEMVALNPLR 78







Query Match	Best Local Similarity	Score 255;	DB 1;	Length 2039;
Matches 67;	Conservative 39;	Mismatches 83;	Indels 32;	Gaps 13;
3 CNE---LPBRRTTELTGWSMDQYTPREGQALYKCRPGYRSLGNVIMCRKKEWALNPL	21.9%;	DB 1;	Length 2039;	59
43 CNAPEWLPFARTNLT---DEFFPPIGYLNEYCRPGSGRPFSSITCLKNSWVGAKD-	30.3%;	Pred. No. 1.4e-14;		97
60 KCKQKRPQGHGDFPFQFTLLTGNAVFEAGVAVYTCNEGVLGEINVRG----				115
98 -RCRRKSCRNPDDPVGWVHIK--IQFGQIKRSCKRGVLLIS-SSATCITISGDYI				153
116 WTNDIPTEIVAK-LPVTAPENGKATVSSAMBEDREYHFGQAVRFVCSG-----YKIEG				168

```

Db 154 WDNTEPIODIRPOLGPLET1-TNDFISTNNE---NFHYSVTVYRCMPGSGGRKVELWG 209
Oy 169 DEMHC--SDD--GFWSKREKPKC-VELSCSPDYVINSPTS 204
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 210 EPSTYCTSNDOYGVISGWPAPQCIIIPKCTPPNVENGLIWS 250

RESULT 5
LEM2_PIG STANDARD: PRT: 484 AA.
AC P98110;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE E-SELECTIN PEPCRSOR (ENDOTHELIAL LEUKOCYTE ADHESION MOLECULE 1)
DE (ELAM-1) (LEUKOCYTE-ENDOTHELIAL CELL ADHESION MOLECULE 2) (LECAM2)
DE (CD62E).
GN SELE.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-AORTIC ENDOTHELIUM;
RX MEDLINE; 95071392.
RA Rollins S.A., Evans M.J., Johnson K.K., Elliot E.A., Squinto S.P.,
RA Mattis L.A., Rother R.P.;
RT "Molecular and functional analysis of porcine E-selectin reveals a
RT potential role in xenograft rejection.";
RN Biochem. Biophys. Res. Commun. 204:763-771(1994).
[2]
RP SEQUENCE FROM N.A.
RC TISSUE-AORTIC ENDOTHELIUM;
RX MEDLINE; 94271236.
RA Tsang Y.T.M., Haekard D.O., Robinson M.K.;
RT "Cloning and expression kinetics of porcine vascular cell adhesion
RT molecule.";
RL Biochem. Biophys. Res. Commun. 201:805-805(1994).
CC -1- FUNCTION: EXPRESSED ON CYTOKINE INDUCED ENDOTHELIAL CELLS AND
CC MEDIATES THEIR BINDING TO LEUKOCYTES. THE LIGAND RECOGNIZED BY
CC ELAM-1 IS SIMILY-LEWIS X (ALPHA1-3)FUCOSYLATED DERIVATIVES OF
CC POLYISACTOSAMINE THAT ARE FOUND AT THE NONREDUCING TERMINI OF
CC GLYCOLIPIDS).
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN ACUTE CELLULAR ALLOGRAFT
CC REJECTION AND PROBABLY ALSO IN XENOGRAFT REJECTION.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: TO OTHER SELECTINS/LECAMs.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS; PORCINE E-LECTIN LACKS
CC THE HUMAN SUSHI-1 AND -4 EQUIVALENTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L39076; AAA61545.1; -
CC EMBL; U08350; AAA21541.1; -.
CC HSSP; P16581; IESL.
CC INTERPRO; IPR000436; -
CC INTERPRO; IPR000561; -
CC INTERPRO; IPR001304; -
CC INTERPRO; IPR002396; -.
CC PFAM; PF00008; EGF_1.
CC PFAM; PF00059; lectin_c; 1.
CC PFAM; PF00084; sushi; 4.
CC PRINTS; PR00343; SELECTIN.
CC PROSITE; PS00022; EGF_1; 1.
CC PROSITE; PS01186; EGF_2; 1.

```

[illegible]

```

Query Match      21.6% ; Score 251; DB 1; Length 340;
Best Local Similarity 29.6% ; Pred No. 4,3e-15;
Matches 69; Conservative 37; Mismatches 91; Indels 36; Gaps 12.

Oy   1 EDCN---ELPFRNRTEILGWSWSDQTV-PEGTOAIKCRGRGV--SLGNVIMVCKRGEW 53
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   51 EFCNRSCVEPTRLNFASLAKOPYITQWNFPVGTVYEVXRCBGRRELSTKLTLCLQNLIW 110
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Oy   54 VALNPLECKCRKRCRGHGDPRFETFTLLTGNNVEYGVKAIVYCNEXYOLGLGINRECDT 113
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db   111 --STAVEFCRKCKCPNGEIRNGQIDVSGILLE--GATISFSICNCGYKLFEGTS-SLCII 165
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Oy   114 DG----WTNDIPICEVVKCLPYTPAENGKIVSSAMEPDREHYHFHGQAVRFVCSGKKIEED 169
      -|-----|-----|-----|-----|-----|-----|-----|-----|
Db   166 SGSSVQMSDPLPECREIETYPAPQIDNG-IIGKRQ---HYGRQSITYACKKGTTMIGE 221
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Oy   170 EEMHC---DDGWSKERKRC-----VEHSCKSPDVIN-----GSPIRSOK 206
      ::||::|:|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db   222 HSITCTVNDEGEWGGPPCECKRGLSTKSVPYVOKPTTVNVPTTEVSPTSOK 274
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT
DAF_CAVPO ID DAF_CAVPO STANDARD, PRT: 507 AA.
AC Q60401; Q60402; Q60403; Q60404; Q60405; Q60406; P97254; P97255;
AC P97256;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE COMPLEMENT DECAT-ACCELERATING FACTOR PRECURSOR.
GN DAF.
OC Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystriocognathi; Cavidae; Cavia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HARTLEY, TISSUE=SPLEEN;
RX MEDLINE: 95403978.
RA Nonaka M., Miwa T., Okada N., Nonaka M., Okada H.;
RT "Multiple isoforms of guinea pig decay-accelerating factor (DAF)
RT generated by alternative splicing."
RL J. Immunol. 155:3037-3048(1995)
CC -1- FUNCTION: PREVENTS THE FORMATION AND/OR ACCELERATES THE
CC DISSOCIATION OF C3 CONVERTASE.
CC -1- SUBCELLULAR LOCATION: CLASS GPI: ATTACHED TO THE MEMBRANE BY A
CC GPI-ANCHOR.
CC -1- ALTERNATIVE PRODUCTS: THERE ARE AT LEAST SIX FORMS OF DAF
CC TRANSCRIPTS (DISTRIBUTED IN FOUR CLASSES: GPI, TCL, TCS AND SEC)
CC GENERATED BY ALTERNATIVE SPLICING. ONE GPI-ANCHORED FORM (GDAB-
CC GPI), TWO FORMS WITH LONGER CYTOPLASMIC REGION (GDAB-TCL AND
CC GDAB-C-TCL), TWO FORMS WITH SHORTER CYTOPLASMIC REGION (GDA-TCS AND
CC GDAB-TCS) AND ONE SECRETED FORM (GDAB-SEC).
CC -1- TISSUE SPECIFICITY: ALL THE ISOFORMS ARE WIDELY EXPRESSED. GPI AND
CC TCS ARE THE MAJOR FORMS, WHEREAS SEC IS MINOR AND TCL IS ONLY
CC PRESENT IN TRACE LEVELS.
CC -1- SIMILARITY: CONTAINS A SUSHI (SCR) REPEATS.
CC -1- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION
CC (RCA) FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
CC -----
DR EMBL: D55667; BAA09514.1; -.
DR EMBL: D55665; BAA09514.1; JOINED.
DR EMBL: D55657; BAA09514.1; JOINED.
DR EMBL: D55658; BAA09514.1; JOINED.
DR EMBL: D55659; BAA09514.1; JOINED.
DR EMBL: D55660; BAA09514.1; JOINED.

```





```

RP SEQUENCE OF 49-88.
RX MEDLINE: 85296001.
RA Chung L.P., Gagnon J., Reid K.B.M.;
RT "Amino acid sequence studies of human C4b-binding protein: N-terminal
RT sequence analysis and alignment of the fragments produced by limited
RT proteolysis with chymotrypsin and the peptides produced by cyanogen
RT bromide treatment."
RL Mol. Immunol. 22:427-435(1985).
RN 171
RP ELECTION MICROSCOPY, 3-DIMENSIONAL STRUCTURE, AND LIGAND BINDING.
RA MEDLINE: 83221615.
RX Dahlbeck B., Smith C.A., Mueller-Eberhard H.J.;
RT "Vitamin K-dependent protein S and complement protein C4b.":
RT vitamin K-dependent protein S and complement protein C4b.":
RL Proc. Natl. Acad. Sci. U.S.A. 80:3461-3465(1983).
CC -1- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
CC ACTIVATION. IT BINDS AS A COFACTOR TO C3b/C4b INACTIVATOR
CC (C3bINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4b. IT
CC ALSO ACCELERATES THE DEGRADATION OF THE C4bC2a COMPLEX (C3
CC CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2a. ALPHA
CC CHAIN BINDS C4b. IT INTERACTS ALSO WITH ANTICOAGULANT PROTEIN S
CC AND WITH SERUM AMYLOID P COMPONENT.
CC -1- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS
CC OF 3 POSSIBLE SORTS: A 570 KDA COMPLEX OF 7 ALPHA CHAINS AND 1
CC BETA CHAIN, A 530 KDA HOMOPOLYMER OF ALPHA CHAINS OR A 500 KDA
CC COMPLEX OF 6 ALPHA CHAINS AND 1 BETA CHAIN. THE CENTRAL BODY OF
CC THE ALPHA CHAIN HOMOPOLYMER SUPPORTS TENTACLES, EACH WITH THE
CC BINDING SITE FOR C4b AT THE END.
CC -1- TISSUE SPECIFICITY: CHYLOMICRONS IN THE PLASMA.
CC -1- SIMILARITY: CONTAINS 8 SUSHI (SCR) REPEATS.
CC -1- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-17 IS THE INITIATOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M31452; AAA36507.1; -.
DR EMBL: M62486; AAA36506.1; -.
DR EMBL: M62475; AAA36506.1; JOINED.
DR EMBL: M62476; AAA36506.1; JOINED.
DR EMBL: M62477; AAA36506.1; JOINED.
DR EMBL: M62478; AAA36506.1; JOINED.
DR EMBL: M62479; AAA36506.1; JOINED.
DR EMBL: M62480; AAA36506.1; JOINED.
DR EMBL: M62481; AAA36506.1; JOINED.
DR EMBL: M62482; AAA36506.1; JOINED.
DR EMBL: M62484; AAA36506.1; JOINED.
DR EMBL: M62485; AAA36506.1; JOINED.
DR EMBL: X07853; CAA30701.1; -.
DR EMBL: X04284; CAB51244.1; -.
DR EMBL: X04296; CAA27839.1; -.
DR EMBL: X02865; CAA26617.1; -.
DR PIR: A33568; NBHUC4.
DR HSSP: P10998; 1VVC.
DR MIM: 120830; -.
DR INTERPRO: IPR000436; -.
DR PFAM: PF00084; sush1; 8.
KW Complement pathway; Plasma; Glycoprotein; Repeat; Sush1; Signal;
KW Polymorphism.
FT SIGNAL 1 48
FT CHAIN 49 597 C4b-BINDING PROTEIN ALPHA CHAIN.
FT DOMAIN 49 539 8 X SUSHI (SCR) REPEATS.
FT REPEAT 49 109 SUSHI 1.
FT REPEAT 112 171 SUSHI 2.
FT REPEAT 174 235 SUSHI 3.
FT REPEAT 238 295 SUSHI 4.
FT REPEAT 298 361 SUSHI 5.
FT REPEAT 364 423 SUSHI 6.

```

Query Match	Best Local Similarity	Score 244;	DB 1;	Length 597;
Matches 59;	Conservative	34;	Mismatches 77;	Indels 30; Gaps 10.
20	SPQTPVEGQAIYKRCGY-RSLGNIYIMCR-GEWVALNPLRKQRCRPGHGDTPFCT	77		
67	TETREFTGTTLTKYTCLPAGVGRSHSTQTLTCONSDEWV-YNTE--CIYKRCRHGE-----	118		
78	FLTLGGNV-----FEYGAAYVTTCNNGYOLLEINREDDTD-----GWTNDIPICPYKAC	128		
119	--LRNQVVEIKIDLSSQSIIEFSCSGFLPLIGSTTSR-CEVODRGVGSHPILPQCEIYVC	175		
129	LPTAENCKIYSSAMEPREYHFGAVPFCVCSGKICGDEEMHSCSD-----GFWSKER	184		
176	KPPDIRNGHSE-----ENFYAGFSVIYSCDPRSLGHASICTYVENETIGVWRPSP	231		
185	PKCVEISCKSPDYINGSPIS	204		
232	PTCEKITCRKPDVSHGEMWS	251		
RESULT 10				
MCP_HUMAN	ID	STANDARD;	PRT;	377 AA.
AC	P15529;			
DT	01-APR-1990 (Rel. 14, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	MEMBRANE COPACITOR PROTEIN PRECURSOR (CD46 ANTIGEN) (TROPHOBLAST			
DE	LEUCOCYTE COMMON ANTIGEN) (TLX).			
GN	MCP.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
RN	(1)			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 35-58.			
RX	MEDLINE; 88286080.			
RA	Lablin D.M., Liszewski M.K., Post T.W., Arce M.A., Le Beau M.M.,			
RA	Rebetics M.B., Lemons R.S., Seya T., Atkinson J.P.;			
RT	"Molecular cloning and chromosomal localization of human membrane			
RT	cocfactor protein (MCP). Evidence for inclusion in the multigene			
RT	family of complement-regulatory proteins."			

RL J. Exp. Med. 168:181-194(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-TESTIS;  
 RX MEDLINE; 93119658.  
 RA Cervoni F., Fenchel P., Akhoundi C., Hsi B.L., Rossi B.;  
 RT "Characterization of a cDNA clone coding for human testis membrane  
 RT cofactor protein (MCP, CD46).";  
 RL Mol. Reprod. Dev. 34:107-113(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.  
 RX MEDLINE; 91267562.  
 RA Purcell D.F., Russell S.M., Deacon N.J., Brown M.A., Hooker D.J.,  
 RA McKenzie I.F.;  
 RT "Alternatively spliced RNAs encode several isoforms of CD46 (MCP), a  
 RT regulator of complement activation.";  
 RL Immunogenetics 33:335-344(1991).  
 RN [4]  
 RP SEQUENCE OF 1-34 FROM N.A.  
 RX MEDLINE; 94014356.  
 RA Cui W., Hourcade D., Post T., Greenlund A.C., Atkinson J.P.,  
 RA Kumar V.;  
 RT "Characterization of the promoter region of the membrane cofactor  
 RT protein (CD46) gene of the human complement system and comparison to  
 RT a membrane cofactor protein-like genetic element.";  
 RL J. Immunol. 151:4137-4146(1993).  
 RN [5]  
 RP ALTERNATIVE SPLICING.  
 RX MEDLINE; 92289809.  
 RA Russell S.M., Sparrow R.L., McKenzie I.F.C., Purcell D.F.J.;  
 RT "Tissue-specific and allelic expression of the complement regulator  
 RT CD46 is controlled by alternative splicing.";  
 RL Eur. J. Immunol. 22:1513-1518(1992).  
 CC -1- FUNCTION: MCP IS AN IMPORTANT MEMBRANE PROTEIN FOR PROTECTING HOST  
 CC CELLS FROM DAMAGE BY COMPLEMENT. MCP POSSESSES COFACTOR ACTIVITY  
 CC FOR INACTIVATION OF C3B AND C4B BY SERUM FACTOR I. THIS WOULD  
 CC ALLOW MCP TO REGULATE THE COMPLEMENT SYSTEM AT THE LEVEL OF THE C3  
 CC CONVERTASES. MAY BE INVOLVED IN THE FUSION OF THE SPERMATOZOA WITH  
 CC THE OOCYTE DURING FERTILIZATION. MAY ACT BY PREVENTING COMPLEMENT  
 CC ACTIVATION AT THE LATER-MATERNAL INTERFACE ON THE  
 CC SYNCYTOTROPHOBLAST FETAL-PLACENTA.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 14 FORMS ARE PRODUCED BY  
 CC ALTERNATIVE SPLICING. THE FORM SHOWN IS THE TRANSCRIPT A. THE  
 CC GAMMA ISOFORM (TRANSCRIPTS A AND B) IS PREFERENTIALLY EXPRESSED IN  
 CC EBV-B CELLS AND LEUKEMIC CELLS. THE 66 KDA ALPHA ISOFORM AND THE  
 CC 56 KDA BETA ISOFORM FOUND IN ALL TISSUES EXCEPT SPERM CORRESPOND  
 CC TO THE TRANSCRIPTS C AND D. AND THE TRANSCRIPTS E AND F  
 CC RESPECTIVELY. THE EXON 9 IS SPECIFICALLY DELETED IN SOME  
 CC PLACENTAE. THE EPSILON ISOFORM CORRESPONDS TO THE TRANSCRIPTS I  
 CC AND J. SPERMATOZOA DELTA ISOFORM (TRANSCRIPT N) DELETE BOTH EXONS  
 CC 12 AND 13. ALL TISSUES DIFFERENTIALLY SPLICE EXON 13.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED BY ALMOST ALL CELLS EXCEPT  
 CC ERYTHROCYTES AND SOME BONE MARROW CELLS.  
 CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED IN THE SER/THR-RICH DOMAIN  
 CC (PROBABLY).  
 CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS.  
 CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD46 entry;  
 CC WWW="HTTP://WWW.NCBI.NLM.NIH.GOV/PROW/CD/CD46.HTM".  
 CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; Y00651; CAA68675.1; -  
 CC EMBL; S51940; AAB24802.1; -  
 CC EMBL; M58050; AAA62833.1; -  
 CC EMBL; A18585; CAA01400.1; -  
 CC EMBL; S65879; AAD13968.1; -

DR PIR; S01896; S01896.  
 DR HSSP; P10998; 1VVC.  
 DR MIM; 120920; -  
 DR INTERPRO; IPR000436; -  
 DR PFAM; PF00084; sush1; 4.  
 KW Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;  
 KM sush1; Alternative splicing.  
 FT SIGNAL 1 34  
 FT CHAIN 35 377  
 FT DOMAIN 35 328  
 FT TRANSMEM 329 351  
 FT DOMAIN 352 377  
 FT DOMAIN 35 284  
 FT REPEAT 35 95  
 FT REPEAT 98 158  
 FT REPEAT 161 224  
 FT REPEAT 227 284  
 FT DOMAIN 287 311  
 FT DISULFID 35 80  
 FT DISULFID 64 94  
 FT DISULFID 99 141  
 FT DISULFID 127 157  
 FT DISULFID 162 210  
 FT DISULFID 191 223  
 FT DISULFID 228 270  
 FT DISULFID 256 283  
 FT CARBOHYD 83 83  
 FT CARBOHYD 114 114  
 FT CARBOHYD 273 273  
 FT CARBSPLIC 286 300  
 FT VARSPLIC 340 352  
 FT VARSPLIC 353 361  
 FT VARSPLIC 362 377  
 FT VARSPLIC 362 377  
 FT VARSPLIC 340 361  
 FT VARSPLIC 339 339  
 SQ SEQUENCE 377 AA; 42247 MW; 2CA6F61752570B57 CRC64;  
 Query Match 20.5%; Score 238.5; DB 1; Length 377;  
 Best Local Similarity 29.6%; Pred. No. 6.1e-14;  
 Matches 64; Conservative 32; Mismatches 99; Indels 21; Gaps 9;  
 QY 2 DCELEPRRNFTEILTGSWSDQTYPEGTQATYKCRPGY--RSIGNVIMCRKGEWALNP 58  
 DB 33 DACEEPTTFEAMELIGK-PKPYEIGERVDYKCKKGFIPLATHHTICDRNHTLPLVSD 91  
 QY 59 LRCKORPGHPGDTFGFTLLGNAVFEYGAAYTNGCYOLLG-EIYRCDTDG-- 115  
 DB 92 -DACYETCTYIND-PLNGOAVPANGTYERGOMHPLCNGYLLIEELIY--CEKGSV 147  
 QY 116 -WTNDPICEVVKCLPYTAPENGIKIVSSAMEPDREYHFGOAVFVNSG----YKIEG 168  
 DB 148 AIWSGKPLICEKVLCTPPRIKNGKHDFEVE---VFELYDNTYSCDPAFGDPISLIG 204  
 QY 169 DEEMHSDGDFWSEKPKVEISKSPDYVINGSPIS 204  
 DB 205 ESTIYGDNWSWRAPECKVYCRPPVYENGKQIS 240  
 RESULT 11  
 CCPH\_HSVSA  
 ID CCPH\_HSVSA STANDARD; PRT; 360 AA.  
 AC 001016;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE COMPLEMENT CONTROL PROTEIN HOMOLOG PRECURSOR (CCPH).  
 GN 4 OR CCPH.  
 OS Herpesvirus saimiri (strain 11).

CC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 CC Gammathepesvirinae; Rhadinovirus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE: 92333688.  
 RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,  
 RA Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B.,  
 RA Honess R.W.,  
 RA "Primary structure of the herpesvirus saimiri genome.";  
 RA J. Virol. 66:5047-5058(1992).  
 RN [2]  
 RP SIMILARITY TO CCP.  
 RA MEDLINE: 92260674.  
 RA Albrecht J.-C., Fleckenstein B.;  
 RT "New member of the multigene family of complement control proteins in  
 RT herpesvirus saimiri.";  
 RL J. Virol. 66:3937-3940(1992).  
 CC -1- ALTERNATIVE PRODUCTS: A MEMBRANE-BOUND FORM AND A SECRETED FORM  
 CC ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.  
 CC -1- SIMILARITY: BELONGS TO THE SUPERFAMILY OF THE REGULATORS OF  
 CC COMPLEMENT ACTIVATION (RCA).  
 CC -1- SIMILARITY: CONTAINS 3 SUSHI (SCR) REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X64346; CAA45626.1; -;  
 DR EMBL: X64346; CAA45627.1; -;  
 DR EMBL: X60283; CAA42823.1; -;  
 DR EMBL: X60283; CAA42822.1; -;  
 DR PIR: B42534; WMBE2E.  
 DR PIR: A42534; WMBE1E.  
 DR PIR: S24567; S24567.  
 DR HSSP: P10998; IYVC.  
 DR INTERPRO: IPR000436; -;  
 DR PFAM: PF00084; sush1; 4.  
 KN Signal: Repeat; Sush1; Transmembrane; Alternative splicing;  
 KN Glycoprotein.  
 FT SIGNAL 1 20 POTENTIAL.  
 FT CHAIN 21 360 COMPLEMENT CONTROL PROTEIN HOMOLOG.  
 FT DOMAIN 83 265 3 X SUSHI (SCR) REPEATS.  
 FT REPEAT 83 143 SUSHI 1.  
 FT REPEAT 146 206 SUSHI 2.  
 FT REPEAT 209 265 SUSHI 3.  
 FT TRANSMEM 328 350 POTENTIAL.  
 FT DISULFID 84 125 BY SIMILARITY.  
 FT DISULFID 111 142 BY SIMILARITY.  
 FT DISULFID 147 191 BY SIMILARITY.  
 FT DISULFID 175 205 BY SIMILARITY.  
 FT DISULFID 210 252 BY SIMILARITY.  
 FT DISULFID 238 264 BY SIMILARITY.  
 FT CARBOHYD 36 36 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 39 39 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 155 155 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 294 294 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 289 302 R1CNGNCTTSMP10 -> AECAPGSNTP155 (IN  
 FT VARSPLIC 303 360 SHORT ISOFORM).  
 FT VARSPLIC 360 40006 MW: 6278A6C2ECD49669 CRC64;  
 SO SEQUENCE 360 AA: 40006 MW: 6278A6C2ECD49669 CRC64;  
 QY Query Match 20.3%; Score 236; DB 1; Length 360;  
 Best Local Similarity 33.3%; Pred. No. 9, 6e-14;  
 Matches 61; Conservative 17; Mismatches 89; Indels 16; Gaps 7;  
 24 YPGSTAIYKCRPGYSLGVIWVCRKGEWVALPLKCKQKRCGHPGDPFGFTLTG 83

DB 44 YNGTTLHTYTCDEGAYKAPVQVITCVNGMWY---PKKQKKKCSPPDLLNGRIYTV-G 99  
 QY 84 NVEYGVKAVTCNNEGYYQLLEIN---YRECDTGDWINDIPICEVYKCLPYAPENGKI 139  
 DB 100 NLY-VGSVITTCNSGYSILGISTTSACLLKRGGRVWTPRPPICDIKKCPPOIANG-- 156  
 QY 140 VSSAMPEDEYHFGCAVRVCNSGK--IEGDENHCSDGFW-SKEPKCVIEISCKSPD 196  
 DB 157 --THTNVKDFYTYLDITVYSCNDETKLTLTGSSKLCSEFGSWPNGETKCFIPLQ 214  
 QY 197 VIN 199  
 DB 215 VAN 217  
 RESULT 12  
 ID DAFL\_MOUSE STANDARD: PRT: 390 AA.  
 AC 061475; 061397; P97732;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE COMPLEMENT DECAY-ACCELERATING FACTOR, GPI-ANCHORED PRECURSOR  
 DE (DAF-GPI).  
 GN DAFL.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=TESTIS;  
 RX MEDLINE: 95403982;  
 RA Spicer A.P., Seidlin M.F., Gendler S.J.;  
 RT "Molecular cloning and chromosomal localization of the mouse decay-  
 RT accelerating factor genes. Duplicated genes encode  
 RT glycosylphosphatidylinositol-anchored and transmembrane forms.";  
 RL J. Immunol. 155:3079-3091(1995).  
 RN [2]  
 RP SEQUENCE OF 7-390 FROM N.A.  
 RC STRAIN=BALB/C; TISSUE=SPLEEN;  
 RX MEDLINE: 96362213.  
 RA Fukuko Y., Yasui A., Okada N., Okada H.;  
 RT "Molecular cloning of murine decay accelerating factor by  
 RT immunoscreening.";  
 RL Int. Immunol. 8:379-385(1996).  
 CC -1- FUNCTION: PROTECTION OF CELLS FROM COMPLEMENT-MEDIATED DAMAGE (BY  
 CC SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.  
 CC -1- TISSUE SPECIFICITY: BRAIN, SECRETORY EPITHELIA, SKELETAL MUSCLE,  
 CC LIVER, TESTES, THYMUS, SPLEEN AND LYMPH NODE.  
 CC -1- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR THE  
 CC FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE  
 CC ACTIVE SITE ON SCR3 (BY SIMILARITY).  
 CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS.  
 CC -1- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION  
 CC (RCA) FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: I41366; AAB00091.1; -;  
 DR EMBL: D63679; BAA09830.1; -;  
 DR HSSP: P08603; 1HCC.  
 DR MCD: MGI:104850; DAFL.  
 DR INTERPRO: IPR000436; -;  
 DR PFAM: PF00084; sush1; 4.  
 KW Complement pathway; Glycoprotein; Membrane; Repeat; GPI-anchor;

KW Signal: Sushi. 1 34  
 FT SIGNAL 35 361  
 FT CHAIN 35 361  
 FT PROPEP 362 390  
 FT DOMAIN 35 285  
 FT REPEAT 35 95  
 FT REPEAT 97 159  
 FT REPEAT 162 221  
 FT REPEAT 224 285  
 FT DOMAIN 288 364  
 FT DISULFID 65 94  
 FT DISULFID 98 145  
 FT DISULFID 129 158  
 FT DISULFID 163 204  
 FT DISULFID 190 220  
 FT DISULFID 225 267  
 FT DISULFID 253 284  
 FT CARBOHYD 187 187  
 FT CARBOHYD 262 262  
 FT LIPID 361 361  
 FT CONFLICT 9 7  
 FT CONFLICT 83 9  
 FT CONFLICT 91 91  
 FT CONFLICT 135 135  
 FT CONFLICT 173 173  
 FT CONFLICT 180 180  
 SQ SEQUENCE 390 AA; 42618 MW; 4418721DFF47F8E7 CRC64;

POTENTIAL.  
 COMPLEMENT DEGRADATION ACCELERATING FACTOR,  
 GPI-ANCHORED.  
 REMOVED IN MATURE FORM (BY SIMILARITY).  
 4 X SUSHI (SCR) REPEATS.  
 SUSHI 1.  
 SUSHI 2.  
 SUSHI 3.  
 SUSHI 4.  
 SER/THR-RICH.  
 BY SIMILARITY.  
 BY SIMILARITY.  
 BY SIMILARITY.  
 BY SIMILARITY.  
 BY SIMILARITY.  
 BY SIMILARITY.  
 BY SIMILARITY.  
 N-LINKED (GLCNAC... ) (POTENTIAL).  
 GPI-ANCHOR (BY SIMILARITY).  
 P -> A (IN REF. 2).  
 T -> A (IN REF. 2).  
 E -> G (IN REF. 2).  
 E -> G (IN REF. 2).  
 E -> K (IN REF. 2).  
 H -> L (IN REF. 2).  
 I -> T (IN REF. 2).

Query Match 19.9%; Score 231.5; DB 1; Length 390;  
 Best Local Similarity 25.2%; Pred. No. 2.6e-13;  
 Matches 64; Conservative 42; Mismatches 81; Indels 67; Gaps 13;  
 QY 3 CN-----ELPPRRNTL--TGSWSDQT-----YPESTQ 29  
 DB 65 CNGGFKGVPPKSNIVCLNGQMSHETFCCKSCVAPRLSFASLUKKEYLMMNFPVGTI 124  
 QY 30 AIKCRGYSLSGVINWCKRGVVALNPL-----KKCKRPGCHGDPFPGFTTLTG 83  
 DB 125 VERECRGEFEDQPL-----PGKATCLEDLWSPVAFQCKKSCPNRPDLNGHINIP 179  
 QY 84 NVEYGVKAVYTCNEGYYQLGEINRECDTIDG---WTNDIPICEVVKCLPVTAPENGKI 139  
 DB 180 ILF--GSEINFSNCPGVRIRIG-VSSTFCSTGNTVMDDEPVTETLHC-----PEPPKI 231  
 QY 140 VSSAMEPDR-YHFGQAVRFVCSNGYKIEGDEMHG---SDGFWSEKPKVE---IS 191  
 DB 232 NNCIMRGESDSYYSQVYVSCDKGFLVGNASIVSKSDVGQWSSPPRCIEKSKVP 291  
 QY 192 CRSPDV---INGSP 202  
 DB 292 TKRPJTIVPSTGTP 305

RESULT 13  
 LEM2\_HUMAN  
 ID LEM2\_HUMAN STANDARD; PRT; 610 AA.  
 AC P16581; P16111.  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE E-SELECTIN PRECURSOR (ENDOTHELIAL LEUKOCYTE ADHESION MOLECULE 1)  
 DE (ELAM-1) (LEUKOCYTE-ENDOTHELIAL CELL ADHESION MOLECULE 2) (LECAM2)  
 DE (CD62E)  
 GN SELE OR ELAM1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 90175359.

RA Hession C., Osborn L., Goff D., Chi-Rosso G., Vassallo C.,  
 RA Pasek M., Pittack C., Tizard R., Goetz S., McCarthy K., Hople S.,  
 RA Lobb R.;  
 RT "Endothelial leukocyte adhesion molecule 1: direct expression cloning  
 RT and functional interactions.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:1673-1677(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 89162047.  
 RA Bevilacqua M.P., Stengelin S., Gimbrone M.A. Jr., Seed B.;  
 RT "Endothelial leukocyte adhesion molecule 1: an inducible receptor for  
 RT neutrophils related to complement regulatory proteins and lectins.";  
 RL Science 243:1160-1165(1989).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 9115870.  
 RA Collins T., Williams A., Johnston G.I., Kim J., Eddy R., Shows T.,  
 RA Gimbrone M.A. Jr., Bevilacqua M.P.;  
 RT "Structure and chromosomal location of the gene for endothelial-  
 RT leukocyte adhesion molecule 1.";  
 RL J. Biol. Chem. 266:2466-2473(1991).  
 RN [4]  
 RP LIGAND.  
 RX MEDLINE; 91068005.  
 RA Phillips M.L., Nudelmann E., Gaeta F.C., Perez M., Singhal A.K.,  
 RA Hakomori S., Paulson J.C.;  
 RT "ELAM-1 mediates cell adhesion by recognition of a carbohydrate  
 RT ligand, sialyl-Lex.";  
 RL Science 250:1130-1132(1990).  
 RN [5]  
 RP 3D-STRUCTURE MODELING OF LECTIN DOMAIN.  
 RX MEDLINE; 93202275.  
 RA Mills A.;  
 RT "Modelling the carbohydrate recognition domain of human E-selectin.";  
 RL FEBS Lett. 319:5-11(1993).  
 RN [6]  
 RP 3D-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 22-178.  
 RX MEDLINE; 94150646.  
 RA Graves B.J., Crowther R.L., Chandran C., Rumberger J.M., Li S.,  
 RA Huang K.-S., Presky D.H., Familletti P.C., Wolitzky B.A., Burns D.K.;  
 RT "Insight into E-selectin/ligand interaction from the crystal  
 RT structure and mutagenesis of the Lec/EGF domains.";  
 RL Nature 367:532-538(1994).  
 RN [7]  
 RP VARIANT ARG-149.  
 RX MEDLINE; 95179107.  
 RA Wenzel K., Felix S., Kleber F.X., Brachold R., Wenke T., Schattke S.,  
 RA Schulte K.L., Glaser C., Rohde K., Baumann G., Speer A.;  
 RT "E-selectin polymorphism and atherosclerosis: an association study.";  
 RL Hum. Mol. Genet. 3:11935-11937(1994).  
 RN [8]  
 RP VARIANT ARG-149.  
 RX MEDLINE; 99134508.  
 RA Ye S.O., Usher D., Virgil D., Zhang L.Q., Yochim S.E., Gupta R.;  
 RT "A patl polymorphism detects the mutation of serine-128 to arginine in  
 RT CD 62E gene--a risk factor for coronary artery disease.";  
 RL J. Biomed. Sci. 6:18-21(1999).  
 CC -1- FUNCTION: EXPRESSED ON CYTOKINE INDUCED ENDOTHELIAL CELLS AND  
 CC MEDIATES THEIR BINDING TO LEUKOCYTES. THE LIGAND RECOGNIZED BY  
 CC ELAM-1 IS SIALYL-Lewis X (ALPHA(1->3)FUCOSYLATED DERIVATIVES OF  
 CC POLYLACTOSAMINE THAT ARE FOUND AT THE NONREDUCING TERMINI OF  
 CC GLYCOPOLYIDS).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- POLYMORPHISM: A POLYMORPHISM IN POSITION 149 IS ASSOCIATED WITH A  
 CC HIGHER RISK OF CORONARY ARTERY DISEASE (CAD). A SIGNIFICANTLY  
 CC HIGHER MUTATION FREQUENCY (ARG-149) IS OBSERVED IN PATIENTS WITH  
 CC ANGIOGRAPHICALLY PROVEN SEVERE ATHEROSCLEROSIS COMPARED WITH AN  
 CC UNSELECTED POPULATION (SER-149).  
 CC -1- SIMILARITY: TO OTHER SELECTINS/LECAMS.  
 CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 6 SUSHI (SCR) REPEATS.  
 CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD62E entry;

```

FT CARBOHYD 179 179 N-LINKED (GLCNAC . .) (POTENTIAL) .
FT CARBOHYD 199 199 N-LINKED (GLCNAC . .) (POTENTIAL) .
FT CARBOHYD 203 203 N-LINKED (GLCNAC . .) (POTENTIAL) .
FT CARBOHYD 265 265 N-LINKED (GLCNAC . .) (POTENTIAL) .
FT CARBOHYD 312 312 N-LINKED (GLCNAC . .) (POTENTIAL) .
FT CARBOHYD 332 332 N-LINKED (GLCNAC . .) (POTENTIAL) .
FT CARBOHYD 503 503 N-LINKED (GLCNAC . .) (POTENTIAL) .
FT CARBOHYD 527 527 N-LINKED (GLCNAC . .) (POTENTIAL) .
FT VARIANT 149 149 S -> R (ASSOCIATED WITH A RISK FACTOR FOR CAD).
FT /FTID=VAR_004191.
FT CONFLICT 468 468 H -> Y (IN REF. 2).
FT SO SEQUENCE 610 AA: 66655 MW; 7D43EE3C0D1229229 CRC64;

Query Match 19.8% Score 230.5 DB 1: Length 610;
Best Local Similarity 27.5%; Pred. No. 5,2e-13;
Matches 53: Conservative 30; Mismatches 81; Indels 29; Gaps 10.

OY 23 TYPEGTOALIKCRPGYSRLGNVIMVC-RKGEVVALLPLRKC---CQ--KRPE-----CGHP 70
DB 261 SEPMMWTTCTCFDEEGFELMGAQLQTSSTSGMNDNEPKTKAVTCRAVRQPQGSVNGSH- 319
OY 71 GDTPGTFTLTIGNVFNFGVKAVYICNCSYQLLGEINTRCDTDC-WINDPICEYVKCL 129
DB 320 --SPAGEFL-----FKSSCNFCFECEGFMLQGAPOV-ECTTGQMTQIIVCEAFQCT 368
OY 130 PYTAENGCKIVASMEPDREHYHGGAVRVCSGSKIEDDEMHCSDSGEFWSEKPKCYE 189
DB 369 ALSNERG-YNMCLPSASGSRFYSGSCERFSCHQGYLVKSRILOCPTEGMENEPTECA 427
OY 190 ISCKSPDVINGSP 202
DB 428 VRC---DAVHOFP 437

RESULT 14
C4BP_MOUSE STANDARD: PRT: 469 AA.
AC P08607;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CAB-BINDING PROTEIN PRECURSOR (C4BP).
GN C4BP OR C4BP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A. MEDLINE; 88024997.
RA Kristensen T., Ogata R.T., Chung L.P., Reid K.B.M., Tack B.F.;
RT "cDNA structure of murine C4b-binding protein, a regulatory component
of the serum complement system.";
RL Biochemistry 26:4668-4674(1987).
CC -1- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
ACTIVATION. IT BINDS AS A CORACTOR TO C3B/C4B INACTIVATOR
(C3BIAN), WHICH THEN HYDROLIZES THE COMPLEMENT FRAGMENT C4B. IT
ALSO ACCELERATES THE DEGRADATION OF THE C4BC2A COMPLEX (C3
CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA
CHAIN BINDS C4B. IT INTERACTS ALSO WITH SERUM AMYLOID P COMPONENT.
-1- SUBUNIT: HOMODEPTAMER; NOT COVALENTLY LINKED. MOUSE LACKS THE
BETA CHAIN OF C4BP.
CC -1- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.
CC -1- SIMILARITY: CONTAINS 6 SUSHI (SCR) REPEATS.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-44 IS THE INITIATOR.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M17122; AAA37312.1; ALT_INIT.
DR PIR: A27117; NBMSCA.
DR HSSP: P10998; 1YVC.
DR MGD: MGI:88229; CAPP.
DR INTERPRO: IPR000436; -.
DR PFAM: PF00084; sush1; 6.
KW Complement pathway; Plasma; Glycoprotein; Repeat; Sush1; Signal.
FT SIGNAL 1 56
FT CHAIN 57 469 CAB-BINDING PROTEIN.
FT DOMAIN 57 414 6 X SUSHI (SCR) REPEATS.
FT REPEAT 57 116 SUSHI 1.
FT REPEAT 119 177 SUSHI 2.
FT REPEAT 180 241 SUSHI 3.
FT REPEAT 244 300 SUSHI 4.
FT REPEAT 302 356 SUSHI 5.
FT REPEAT 358 414 SUSHI 6.
FT DISULFID 58 103 BY SIMILARITY.
FT DISULFID 88 115 BY SIMILARITY.
FT DISULFID 120 160 BY SIMILARITY.
FT DISULFID 146 176 BY SIMILARITY.
FT DISULFID 181 223 BY SIMILARITY.
FT DISULFID 209 240 BY SIMILARITY.
FT DISULFID 245 287 BY SIMILARITY.
FT DISULFID 273 299 BY SIMILARITY.
FT DISULFID 303 343 BY SIMILARITY.
FT DISULFID 329 355 BY SIMILARITY.
FT DISULFID 359 400 BY SIMILARITY.
FT DISULFID 386 413 BY SIMILARITY.
FT CARBOHYD 74 74 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 275 275 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 292 292 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 366 366 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 381 381 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 428 428 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 469 AA; 51551 MW; 41E137CB8D8C6321 CRC64;

Query Match 19.6%; Score 228; DB 1; Length 469;
Best Local Similarity 28.1%; Pred. No. 6,5e-13;
Matches 59; Conservative 40; Mismatches 87; Indels 24; Gaps 11;

QY 7 PPRRTTELITGSWSDQT-YPEGTQATYKCRGQY-RSLGNVIMVCR-KGEW-VALNPLKRC 62
DB 60 PPAPALNALPASDVNRTPFESHHTLTKYECPLPGYGRISRMVYCKPSEMEWTSVS----C 115
QY 63 QKPCGHPGDPFGFTLTGNGVFEYGVKAVYTCNEGYYQLGEINVRREGDTDG---WTN 118
DB 116 AKKHCRNGCIYDNG---YVNGETIFFGSOIEFSCQEGFIVNS-STSCFVNRKGVAMSN 171
QY 119 DIPICEVVKCLPVTAPENGKIVSSAMEPDRREYHFGQAVRFVQNSGYKLTGDEEMHCS--- 175
DB 172 PEPCEVIYKCGPPDISNGK--HSGTEDEFPYFNHG--ISTYCDPGRFVSGSPFICTVYN 227
QY 176 -DDGFWSKKEKCVISCKSPVINGSPIS 204
DB 228 KTVPYWSSSPTECKIKISQPNILHGIVIS 257

RESULT 15
LEM3_MOUSE STANDARD: PRT; 768 AA.
AC 001102;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE P-SELECTIN PROTEIN (GRANULE MEMBRANE PROTEIN 140) (GMP-140) (PADGEM)
DE (CD62E) (LEUCOCYTE-ENDOTHELIAL CELL ADHESION MOLECULE 3) (LECAM).
GN SELP OR GRMP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92340571.
RA Weller A., Isenmann S., Vestweber D.;
RT "Cloning of the mouse endothelial selectin. Expression of both E-
RT and P-selectin is inducible by tumor necrosis factor alpha.";
RL J. Biol. Chem. 267:15176-15183(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92345617.
RA Sanders W.E. Jr., Wilson R.W., Ballantyne C.M., Beaudet A.L.;
RT "Molecular cloning and analysis of in vivo expression of murine P-
RT selectin.";
RL Blood 80:795-800(1992).
CC -1- FUNCTION: CA(2+)-DEPENDENT RECEPTOR FOR MYELOID CELLS THAT BINDS
CC TO CARBOHYDRATES ON NEUTROPHILS AND MONOCYTES. MEDIATES THE
CC INTERACTION OF ACTIVATED ENDOTHELIAL CELLS OR PLATELETS WITH
CC LEUKOCYTES. THE LIGAND RECOGNIZED IS STALYV-LEWIS X.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: STORED IN THE ALPHA-GRANULES OF PLATELETS
CC AND WEIBEL-PALADE BODIES OF ENDOTHELIAL CELLS. UPON CELL
CC ACTIVATION BY AGONISTS, P-SELECTIN IS TRANSPORTED RAPIDLY TO
CC THE CELL SURFACE.
CC -1- INDUCTION: BY TUMOR NECROSIS FACTOR ALPHA.
CC -1- SIMILARITY: TO OTHER SELECTINS/LECAMs.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 8 SUSHI (SCR) REPEATS; MOUSE P-LECTIN LACKS
CC THE HUMAN SUSHI-2 EQUIVALENT.
CC -----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M87861; AAA40008.1; -.
DR EMBL: M72332; AAA37712.1; -.
DR PIR: A42755; A42755.
DR HSSP: P16109; 1FSB.
DR MGD: MGI:98280; SELP.
DR INTERPRO: IPR000436; -.
DR INTERPRO: IPR000561; -.
DR INTERPRO: IPR001304; -.
DR INTERPRO: IPR002396; -.
DR PFAM: PF00008; EGF_1.
DR PFAM: PF00059; lectin_c; 1.
DR PFAM: PF00084; sush1; 8.
DR PRINTS: PR00343; SELECTIN.
DR PROSITE: PS00022; EGF_1; 1.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE: PS00615; C-TYPE_LECTIN_2; 1.
DR PROSITE: PS00615; C-TYPE_LECTIN_3; 1.
KW Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
KW Selectin; Signal; Sush1; Repeat; Lipoprotein; Palmitate.
FT SIGNAL 1 41
FT CHAIN 42 768
FT DOMAIN 42 709 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 710 733 POTENTIAL.
FT DOMAIN 734 768 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 158 158 C-TYPE LECTIN (SHORT FORM).
FT DOMAIN 159 195 EGF-LIKE.
FT DOMAIN 199 700 8 X SUSHI (SCR) REPEATS.
FT REPEAT 199 258 SUSHI 1.
FT REPEAT 261 320 SUSHI 2.
FT REPEAT 323 382 SUSHI 3.
FT REPEAT 385 444 SUSHI 4.
FT REPEAT 447 506 SUSHI 5.
FT REPEAT 509 568 SUSHI 6.
FT REPEAT 579 638 SUSHI 7.

```

```
FT REPEAT 641 700 SUSHI 8.  
FT DISULFID 60 158 BY SIMILARITY.  
FT DISULFID 131 150 BY SIMILARITY.  
FT DISULFID 163 174 BY SIMILARITY.  
FT DISULFID 168 183 BY SIMILARITY.  
FT DISULFID 185 194 BY SIMILARITY.  
FT DISULFID 200 244 BY SIMILARITY.  
FT DISULFID 230 257 BY SIMILARITY.  
FT DISULFID 262 306 BY SIMILARITY.  
FT DISULFID 292 319 BY SIMILARITY.  
FT DISULFID 324 368 BY SIMILARITY.  
FT DISULFID 354 381 BY SIMILARITY.  
FT DISULFID 386 430 BY SIMILARITY.  
FT DISULFID 416 443 BY SIMILARITY.  
FT DISULFID 448 492 BY SIMILARITY.  
FT DISULFID 478 505 BY SIMILARITY.  
FT DISULFID 510 554 BY SIMILARITY.  
FT DISULFID 540 567 BY SIMILARITY.  
FT DISULFID 580 624 BY SIMILARITY.  
FT DISULFID 610 637 BY SIMILARITY.  
FT DISULFID 642 686 BY SIMILARITY.  
FT DISULFID 672 699 BY SIMILARITY.  
FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 654 654 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 661 661 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 679 679 PALMITATE (BY SIMILARITY).  
FT LIPID 745 745 ENDOCYTOSIS SIGNAL (PROBABLE).  
FT SITE 756 759 A -> E (IN REF. 2).  
FT CONFLICT 724 724  
SQ SEQUENCE 768 AA; 83098 MW; E5173074D2F66E68 CRC64;
```

```
Query Match 19.6%; Score 227.5; DB 1; Length 768;  
Best Local Similarity 26.4%; Pred. No. 1.2e-12;  
Matches 61; Conservative 41; Mismatches 74; Indels 55; Gaps 13;
```

```
OY 1 EDCNELPRRNTTEITGSMDO-----TYPEGTQAIYKCRPG--YRSLGNVIMCR 49  
Db 129 QDCEVEIYIKSNS--APGKWNDEPCFKRRRALCYTASCQDMSCSNOGECIETIGSYTCSCY 186  
OY 50 KG-----EWVALNPLRRCOK-----RPGGHP-GDTPFGTFTLTGNGVFEYGVKAVY 94  
Db 187 PGFYGPECEYV-----KECGKVINPQHVLAMNSHPLGE-----FSFNSQCTF 228  
OY 95 TCNEGYQL-LGEINVRCDTIDG-WFNNDIPICEVVKCLPVTAPENGKIYSSAMEPDRHYH 151  
Db 229 SCAREGYELDGPGL---QCLASGIWTNPNPKCDAVQCOSLEAPPHGTM--ACMHPIAAFA 283  
OY 152 FGQAVRFYCNAGSYKTEGDEMHGSDGDFWFSKPKCVELISCKSPDV-INGS 201  
Db 284 YDSSCKFECCPGYRARGSNLTLCGTSGQWSEPLPTCEALACEPPEIPIHGS 334
```

```
Search completed: November 21, 2000, 16:49:19  
Job time: 322 sec
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2000, 16:43:22 ; Search time 70.54 Seconds  
(without alignments)  
274.007 Million cell updates/sec

Title: US-09-316-163-9  
Perfect score: 1163  
Sequence: 1 EDCNELPPRRNTTEILGWSMS.....VEISCKSPVINGSPISQKI 207

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 297973 seqs, 93374136 residues

Total number of hits satisfying chosen parameters: 297973

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPREMBL.11:  
1: sp\_apnea:  
2: sp\_bacteria:  
3: sp\_fungi:  
4: sp\_human:  
5: sp\_invertebrate:  
6: sp\_mammal:  
7: sp\_mmc:  
8: sp\_organelle:  
9: sp\_plant:  
10: sp\_rodent:  
11: sp\_virus:  
12: sp\_virus:  
13: sp\_virus:  
14: sp\_unclassified:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1163	100.0	449	4	Q14570
2	499	42.9	669	4	Q28085
3	368.5	31.7	1053	13	Q91275
4	276.5	23.8	645	12	Q9WRU2
5	275	23.6	645	12	Q9YTO8
6	270.5	23.3	550	12	P88903
7	270.5	23.3	550	12	Q40912
8	261	22.4	347	4	P78361
9	260	22.4	679	11	Q99254
10	256	22.0	522	6	Q28769
11	255.5	22.0	559	12	P87616
12	255	21.9	2039	4	Q9UQV2
13	255	21.9	2039	4	Q16745
14	255	21.9	2039	4	Q16744
15	253.5	21.8	482	6	Q28982
16	248	21.3	315	6	Q28770
17	247	21.2	363	6	Q02839
18	247	21.2	661	6	Q29531
19	247	21.2	2014	6	Q29530

20	246.5	21.2	383	11	Q9WRT9	09wt19 cavia porce
21	243.5	20.9	497	11	Q63612	063612 rattus norv
22	243.5	20.9	559	11	Q63135	063135 rattus norv
23	243	20.9	222	6	Q19122	Q19122 callimico g
24	243	20.9	560	5	Q22328	Q22328 caenorhabdi
25	240.5	20.7	483	11	Q64735	Q64735 mus musculu
26	239.5	20.6	1911	6	Q29528	Q29528 papio hamad
27	238.5	20.5	349	4	Q15429	Q15429 homo sapien
28	238.5	20.5	378	6	Q62837	Q62837 saginus oe
29	237	20.4	222	6	Q19128	Q19128 pithecia pl
30	237	20.4	1652	5	Q9V109	Q9V109 dirosophila
31	236.5	20.3	1124	5	Q9VYR4	Q9VYR4 dirosophila
32	236	20.3	263	12	Q89061	Q89061 variola vir
33	236	20.3	263	12	Q89859	Q89859 variola vir
34	236	20.3	263	12	Q07033	Q07033 variola vir
35	235	20.2	263	12	Q89076	Q89076 variola vir
36	233.5	20.1	417	11	Q35520	Q35520 rattus norv
37	232	19.9	222	6	Q19125	Q19125 saimiri sci
38	232	19.9	314	6	Q62835	Q62835 saginus oe
39	232	19.9	377	6	Q62838	Q62838 saginus oe
40	231.5	19.9	285	6	Q19126	Q19126 macaca fasc
41	231.5	19.9	285	6	Q19127	Q19127 macaca mula
42	229.5	19.7	336	6	Q62834	Q62834 saginus oe
43	229.5	19.7	369	6	P79138	P79138 cercopithec
44	228.5	19.6	522	6	Q28797	Q28797 pan troglod
45	228	19.6	222	6	Q19124	Q19124 saginus my

## ALIGNMENTS

RESULT	ID	1	PRELIMINARY:	PRT:	449 AA.
Q14570	Q14570	Q14570	Q14570	Q14570	Q14570
AC	Q14570	Q14570	Q14570	Q14570	Q14570
DT	01-NOV-1996	01-NOV-1996	01-NOV-1996	01-NOV-1996	01-NOV-1996
DI	01-NOV-1996	01-NOV-1996	01-NOV-1996	01-NOV-1996	01-NOV-1996
DT	01-MAY-2000	01-MAY-2000	01-MAY-2000	01-MAY-2000	01-MAY-2000
DE	COMPLEMENT FACTOR H PRECURSOR.	COMPLEMENT FACTOR H PRECURSOR.	COMPLEMENT FACTOR H PRECURSOR.	COMPLEMENT FACTOR H PRECURSOR.	COMPLEMENT FACTOR H PRECURSOR.
GN	HF OR CFH.	HF OR CFH.	HF OR CFH.	HF OR CFH.	HF OR CFH.
OS	Homo sapiens (Human).	Homo sapiens (Human).	Homo sapiens (Human).	Homo sapiens (Human).	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
RN	[1]	[1]	[1]	[1]	[1]
RP	SEQUENCE FROM N.A.	SEQUENCE FROM N.A.	SEQUENCE FROM N.A.	SEQUENCE FROM N.A.	SEQUENCE FROM N.A.
RX	MEDLINE: 88134059.	MEDLINE: 88134059.	MEDLINE: 88134059.	MEDLINE: 88134059.	MEDLINE: 88134059.
RA	Ripoche J., Day A.J., Harris T.J.R., Sim R.B.;	Ripoche J., Day A.J., Harris T.J.R., Sim R.B.;	Ripoche J., Day A.J., Harris T.J.R., Sim R.B.;	Ripoche J., Day A.J., Harris T.J.R., Sim R.B.;	Ripoche J., Day A.J., Harris T.J.R., Sim R.B.;
RT	"The complete amino acid sequence of human complement factor H.;"	"The complete amino acid sequence of human complement factor H.;"	"The complete amino acid sequence of human complement factor H.;"	"The complete amino acid sequence of human complement factor H.;"	"The complete amino acid sequence of human complement factor H.;"
RL	Biochem. J. 249:593-602(1988).	Biochem. J. 249:593-602(1988).	Biochem. J. 249:593-602(1988).	Biochem. J. 249:593-602(1988).	Biochem. J. 249:593-602(1988).
RN	[2]	[2]	[2]	[2]	[2]
RP	SEQUENCE OF 226-449 FROM N.A.	SEQUENCE OF 226-449 FROM N.A.	SEQUENCE OF 226-449 FROM N.A.	SEQUENCE OF 226-449 FROM N.A.	SEQUENCE OF 226-449 FROM N.A.
RX	MEDLINE: 86169701.	MEDLINE: 86169701.	MEDLINE: 86169701.	MEDLINE: 86169701.	MEDLINE: 86169701.
RA	Kristensen T., Wetsel R.A., Tack B.F.;	Kristensen T., Wetsel R.A., Tack B.F.;	Kristensen T., Wetsel R.A., Tack B.F.;	Kristensen T., Wetsel R.A., Tack B.F.;	Kristensen T., Wetsel R.A., Tack B.F.;
RT	"Structural analysis of human complement protein H: homology with C4b	"Structural analysis of human complement protein H: homology with C4b	"Structural analysis of human complement protein H: homology with C4b	"Structural analysis of human complement protein H: homology with C4b	"Structural analysis of human complement protein H: homology with C4b
RL	binding protein, beta 2-glycoprotein I, and the Ba fragment of B2.;"	binding protein, beta 2-glycoprotein I, and the Ba fragment of B2.;"	binding protein, beta 2-glycoprotein I, and the Ba fragment of B2.;"	binding protein, beta 2-glycoprotein I, and the Ba fragment of B2.;"	binding protein, beta 2-glycoprotein I, and the Ba fragment of B2.;"
RN	J. Immunol. 136:3407-3411(1986).	J. Immunol. 136:3407-3411(1986).	J. Immunol. 136:3407-3411(1986).	J. Immunol. 136:3407-3411(1986).	J. Immunol. 136:3407-3411(1986).
RP	SEQUENCE OF 1-19 FROM N.A.	SEQUENCE OF 1-19 FROM N.A.	SEQUENCE OF 1-19 FROM N.A.	SEQUENCE OF 1-19 FROM N.A.	SEQUENCE OF 1-19 FROM N.A.
RA	Vik D.P., Williams S.A.;	Vik D.P., Williams S.A.;	Vik D.P., Williams S.A.;	Vik D.P., Williams S.A.;	Vik D.P., Williams S.A.;
RT	Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.	Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.	Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.	Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.	Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
RL	[4]	[4]	[4]	[4]	[4]
RP	SEQUENCE OF 1-9 FROM N.A.	SEQUENCE OF 1-9 FROM N.A.	SEQUENCE OF 1-9 FROM N.A.	SEQUENCE OF 1-9 FROM N.A.	SEQUENCE OF 1-9 FROM N.A.
RA	Dominguez O.;	Dominguez O.;	Dominguez O.;	Dominguez O.;	Dominguez O.;
RT	Thesis (1993), Immunologia, Hospital Trias I Pujol, Spain.	Thesis (1993), Immunologia, Hospital Trias I Pujol, Spain.	Thesis (1993), Immunologia, Hospital Trias I Pujol, Spain.	Thesis (1993), Immunologia, Hospital Trias I Pujol, Spain.	Thesis (1993), Immunologia, Hospital Trias I Pujol, Spain.
RL	EMBL: X07523; CA30403.1; -	EMBL: X07523; CA30403.1; -	EMBL: X07523; CA30403.1; -	EMBL: X07523; CA30403.1; -	EMBL: X07523; CA30403.1; -
DR	EMBL: M12383; AAA52013.1; -	EMBL: M12383; AAA52013.1; -	EMBL: M12383; AAA52013.1; -	EMBL: M12383; AAA52013.1; -	EMBL: M12383; AAA52013.1; -
DR	EMBL: U56979; AAB01987.1; -	EMBL: U56979; AAB01987.1; -	EMBL: U56979; AAB01987.1; -	EMBL: U56979; AAB01987.1; -	EMBL: U56979; AAB01987.1; -
DR	EMBL: Z29665; CAA82763.1; -	EMBL: Z29665; CAA82763.1; -	EMBL: Z29665; CAA82763.1; -	EMBL: Z29665; CAA82763.1; -	EMBL: Z29665; CAA82763.1; -
DR	HSSP: P10998; IYVC.	HSSP: P10998; IYVC.	HSSP: P10998; IYVC.	HSSP: P10998; IYVC.	HSSP: P10998; IYVC.
DR	INTERPRO: IPR000436; -	INTERPRO: IPR000436; -	INTERPRO: IPR000436; -	INTERPRO: IPR000436; -	INTERPRO: IPR000436; -
DR	PRAM: PF00084; sushi; 7.	PRAM: PF00084; sushi; 7.	PRAM: PF00084; sushi; 7.	PRAM: PF00084; sushi; 7.	PRAM: PF00084; sushi; 7.
KW	SIGNAL.	SIGNAL.	SIGNAL.	SIGNAL.	SIGNAL.
FT	SIGNAL	1	18	POTENTIAL.	POTENTIAL.

FT CHAIN 19 449 POTENTIAL.  
SQ SEQUENCE 449 AA; 51033 MW; C2AAD47F155343E3 CRC64;

Query Match 100.0%; Score 1163; DB 4; Length 449;  
Best Local Similarity 100.0%; Pred. No. 4.5e-108;  
Matches 207; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDCELPFRRTTELITGWSMDQTPGCOALYKCRPGYRSIGNVIMVCRKGEWALNPLR 60  
DB 19 EDCELPFRRTTELITGWSMDQTPGCOALYKCRPGYRSIGNVIMVCRKGEWALNPLR 78  
QY 61 KQKRPQGHDPEDPFGFTLLGNVFEYGVAVYTCNNGYOLLGEINRECDTDGWTNDI 120  
DB 79 KQKRPQGHDPEDPFGFTLLGNVFEYGVAVYTCNNGYOLLGEINRECDTDGWTNDI 138  
QY 121 PICGVVCLPVTAPENKIKYSSAMEPREYHFGAVRFVCSGYKIEGDEMHCSDDGFW 180  
DB 139 PICGVVCLPVTAPENKIKYSSAMEPREYHFGAVRFVCSGYKIEGDEMHCSDDGFW 198  
QY 181 SKEKPKVEISCKSPDIVNGSPISOKI 207  
DB 199 SKEKPKVEISCKSPDIVNGSPISOKI 225

## RESULT 2

Q28085 PRELIMINARY; PRT; 669 AA.

AC Q28085;  
ID 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DE CCP MODULES 3-12, WITH PARTS OF CCP 2 AND 13 (FRAGMENT).  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX TISSUE=LIVER;  
RC MEDLINE; 96202005.  
RA Soames C.J., Day A.J., Sim R.B.;  
RT "Prediction from sequence comparisons of residues of factor H involved  
in the interaction with complement component C3b."  
RL Biochem. J. 315:523-531(1996).  
DR EMBL; X98697; CAA67257.1; -.  
DR HSSP; P10998; IYVC.  
DR INTERPRO; IPR000436; -.  
DR PFAM; PF00084; sushi; 11.  
FT NON\_TER 1  
FT NON\_TER 669  
SQ SEQUENCE 669 AA; 75683 MW; D0D9DB30EB747AC2 CRC64;

Query Match 42.9%; Score 499; DB 6; Length 669;  
Best Local Similarity 64.9%; Pred. No. 1.3e-41;  
Matches 85; Conservative 19; Mismatches 27; Indels 0; Gaps 0;

QY 76 GTFTLLGNVFEYGVAVYTCNNGYOLLGEINRECDTDGWTNDIPICEVVKCLPVTAP 135  
DB 3 GSPHLAEGNQFEGKAVVYTCDEGYOMGEAMFREDCTDGTNDIPICEVVKCLPVTAP 62  
QY 136 NGKIYSSAMEPREYHFGAVRFVCSGYKIEGDEMHCSDDGFWSEKPKVEISCKSP 195  
DB 63 NGKIYSSAMEPREYHFGAVRFVCSGYKIEGDEMHCSDDGFWSEKPKVEISCKSP 122  
QY 196 DIVNGSPISOK 206  
DB 123 VILNGQAVLPK 133

RESULT 3  
Q91275

ID Q91275 PRELIMINARY; PRT; 1053 AA.

AC Q91275;  
ID 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DE 01-MAY-2000 (TREMblrel. 13, Last annotation update)  
OS COMPLEMENT REGULATORY PLASMA PROTEIN.  
OC Parabrax nebulifer (barred sand bass).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Acanthopterygii;  
OC Perciformes; Perciiformes; Percoidae; Serranidae; Parabrax.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX TISSUE=LIVER;  
RC MEDLINE; 94318039.  
RA Dahmen A., Kaidoh T., Zipfel P.F., Giggli I.;  
RT "Cloning and characterization of a cDNA representing a putative  
nebulifer";  
RT complement-regulatory plasma protein from barred sand bass (Parabrax  
nebulifer)."  
RL Biochem. J. 301:391-397(1994).  
DR EMBL; U21703; AAA92556.1; -.  
DR HSSP; P08603; IYFH.  
DR INTERPRO; IPR000436; -.  
DR PFAM; PF00084; sushi; 16.  
SQ SEQUENCE 1053 AA; 117597 MW; F27E32C3AD76D5D3 CRC64;

Query Match 31.7%; Score 368.5; DB 13; Length 1053;  
Best Local Similarity 39.6%; Pred. No. 2.4e-28;  
Matches 72; Conservative 25; Mismatches 76; Indels 9; Gaps 4;

QY 21 DQYRPEQTOALYKCRPGYRSIGNVIMVCRKGEWALNPLRKQKRPQGHDPEDPFGFTL 80  
DB 47 EASTPGGRQYRVCGNNGYS--GFEKLVCEGKMETRG--AKCQPRSCGHPGDAQFAFPHL 102  
QY 81 TGGNVFEYGVAVYTCNNGYOLLGEINRECDTDGWTNDIPICEVVKCLPVTAPENKIV 140  
DB 103 AEGNDFEYFGSKVYVTCQKGYOMSRINRYRCVAGMGVVPVCESSQCLPIHVNNOVI 162  
QY 141 SSAMEPREYHFGAVRFVCSGYKIEGDEMHCSDDGFWSEKPKVEISCKSPDIVN 199  
DB 163 GG---PEEATFGNVVAFSCSRSEILIDGSPELCDERGDMSFPVPCAKAITCAIPLEN 218  
QY 200 GS 201  
DB 219 GN 220

## RESULT 4

Q9WRU2 PRELIMINARY; PRT; 645 AA.

AC Q9WRU2;  
ID 01-NOV-1999 (TREMblrel. 12, Created)  
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)  
DE 01-JUN-2000 (TREMblrel. 14, Last annotation update)  
OS COMPLEMENT BINDING PROTEIN.  
OC Macaca mulatta rhadinovirus 17577.  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Gammaherpesvirinae; Rhadinovirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 99174001.  
RA Seales R.P., Bergquam E.P., Axthelm M.K., Wong S.W.;  
RT "Sequence and genomic analysis of a Rhesus macaque rhadinovirus with  
similarity to Kaposi's sarcoma-associated herpesvirus/human  
herpesvirus 8."  
RL J. Virol. 73:3040-3053(1999).  
DR EMBL; AF083501; AAD21332.1; -.  
DR INTERPRO; IPR000436; -.  
DR PFAM; PF00084; sushi; 8.  
SQ SEQUENCE 645 AA; 71526 MW; 93D8D35ABF61EB2 CRC64;

Query Match 23.8%; Score 276.5; DB 12; Length 645;



```

Db 52 RCRSGYTTANINATCTCGGTW--SEPTATCNKSCPNCEIQLNGKVIFFHGQDALYXG 109
QY 90 VKAYVTCNEGYYQLGEINREC-----DTGWTNDIPICEVYKCLPYAPENGKIVSSAM 144
Db 110 ANISYVCEGFEYFLVREYRCMIGASGQAMWSSSPFCEKEK-----HRPKLENDIF 163
QY 145 EPDREYH-FGQAVRFVNCSSGKIKIEDDEMHGSDGCFMSEKPKVEICKSPDYINGSPI 203
Db 164 KPDADYEDNDVAHFECEGNTLVGPHSICAVANNWTWISNMPCELCAGCKFPSTVHGYP 223

RESULT 8
P78361
ID P78361 PRELIMINARY; PRT; 347 AA.
AC P78361
DT 01-MAY-1997 (TREMblrel. 03, Created)
DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)
DE DECAF-ACCELERATION FACTOR (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN HIPPOCAMPUS;
RA Kumar V.B., Hyung C., Nakira R., Walters M., Sasser T., Bernardo A.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: U88576; AAB48622.1; -.
DR HSSP: P08603; IHCC.
DR INTERPRO: IPR000436; -.
DR PFAM: PF00084; sushi; 4.
FT NON_TER 1
SQ SEQUENCE 347 AA; 37847 MW; 7B123F7F6D48962 CRC64;

Query Match 22.4%; Score 261; DB 4; Length 347;
Best Local Similarity 29.9%; Pred. No. 3.3e-18;
Matches 70; Conservative 41; Mismatches 85; Indels 38; Gaps 12;

QY 1 EDN---ELPPRNTELTGTSWSDQY-PEGTQAIYKCRGGR---SLGNVIMCRKGEW 53
Db 58 EFCNRCCEVPTRLNSASLQKOPTTONTFFVGTVEYECRGRREPLSPKLTCLQMLKW 117
QY 54 VALNPLRCKQCRKPGHGDTPFGTFTLTGQNVFEYGVKAYVTCNEGYYQLGEINRECDT 113
Db 118 --STAVEFCRKKSCPMPEIRNGIDVPGILF--GATIFSCNTGKLFGSTS-SICLI 172
QY 114 DG---WTNDIPICEVYKCLPYAPENGKIVSSAMEPDE-YHFGQAVRFVNCSSGKIEG 168
Db 173 SGGSVQWSDPLPCREIYC-----PAPQIDNGIIQGERDHYGRSVIYACNKGFTMIG 227
QY 169 DEEMHC---SDGFMSKEKPKC-----VEISCKSPDIVN-----GSPISOK 206
Db 228 EHSIYCTVANNDEGEMSGPPPECRKGSLTSKVPTVHKPTTVNPTVTSPTSOX 281

RESULT 9
Q99254
ID Q99254 PRELIMINARY; PRT; 679 AA.
AC Q99254
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DE COMPLEMENT RECEPTOR TYPE 2 (CR2) (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RX MEDLINE: 90229754.
RA Kurtz C.B., O'Toole E., Christensen S.M., Weis J.H.;

```

```

RT "The murine complement receptor gene family. IV. Alternative splicing
RT of Cr2 gene transcripts predicts two distinct gene products that share
RT homologous domains with both human CR2 and CR1."
RL J. Immunol. 144:3581-3591(1990).
RN [2]
RP SEQUENCE OF 21-367 FROM N.A.
RX MEDLINE: 95105691.
RA Kim Y.U., Kinoshita T., Molina H., Hourcade D., Seya T., Wagner L.M.,
RA Holers V.M.;
RT "Mouse complement regulatory protein Cr2/p65 uses the specific
RT mechanisms of both human decay-accelerating factor and membrane
RT cofactor protein."
RL J. Exp. Med. 181:151-159(1995).
DR EMBL: U17128; AAA78271.1; -.
DR EMBL: U17123; AAA78271.1; JOINED.
DR EMBL: U17124; AAA78271.1; JOINED.
DR EMBL: U17125; AAA78271.1; JOINED.
DR EMBL: U17126; AAA78271.1; JOINED.
DR EMBL: U17127; AAA78271.1; JOINED.
DR HSSP: M36470; AAA37449.1; -.
DR MGD: MGI:88489; Cr2.
DR INTERPRO: IPR000436; -.
DR PFAM: PF00084; sushi; 10.
FT NON_TER 1
SQ SEQUENCE 679 AA; 74916 MW; 52FC00FDCED20CDC CRC64;

Query Match 22.4%; Score 260; DB 11; Length 679;
Best Local Similarity 28.4%; Pred. No. 9.5e-18;
Matches 56; Conservative 45; Mismatches 72; Indels 24; Gaps 9;

QY 20 SDQT-YPEGTQAIYKCRPGYRSLGNVIMCRKGEVALNPLRCKQCRKPGHGDTPPGTF 78
Db 37 SDKSEFAIGFTWEYKCRPGYRFRKSFITTCLETJSKW--SDAQPFCKRCPANNPQELHGSV 94
QY 79 TLNGGNVFEYGVKAYVTCNEGYYQLGEINREC-----DTGWTNDIPICEVYKCLPYAP 134
Db 95 HINTG--IEFGSTITTYSCNCGYRLIGD--SSATCIVSDNTVMWMDMLCESIPCESPAI 151
QY 135 ENGIYSSAMEPDEYHFGQAVRFVNCSSG-----YKIEGDEEMHGSDD-----GFWSEK 164
Db 152 SNQDFFYSSRD-----SFFYGNVVTYICHTGKNRKLPLVSEKSYICTSKDNOYGIWNSPP 208
QY 185 PKCY-ELISCKSPDIVNG 200
Db 209 POCIPRYKCPMEIENG 225

RESULT 10
Q28769
ID Q28769 PRELIMINARY; PRT; 522 AA.
AC Q28769
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DE COMPLEMENT RECEPTOR (FRAGMENT).
OS Papio hamadryas (Hamadryas baboon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Papio.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BONE MARROW;
RA Birmingham D.J., Logar C.M., Shen X.P., Chen W.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: L77977; AAA99004.1; -.
DR HSSP: P08603; IHFI.
DR INTERPRO: IPR000436; -.
DR PFAM: PF00084; sushi; 7.
FT NON_TER 1
SQ SEQUENCE 522 AA; 56626 MW; 312FCBE03ADF19DC CRC64;

```

Query Match 22.0%; Score 256; DB 6; Length 522;  
 Best Local Similarity 31.1%; Pred. No. 1.7e-17;  
 Matches 69; Conservative 38; Mismatches 81; Indels 34; Gaps 14;

3 CNE---ELPPRRNTEILGWSMDQTYEGTQAIYKCRGYRSLGNVIMVCRKGEWALNPL 59  
 48 CNAPEQLPFPARPTNLTAS-----EPYGYLYKYECLPGYKGRKPSIICLNKSVWTSKAD- 102  
 60 RKCQRKPCGHPDTPPEGTFTLLTGAVNEFYGVKAVYTCNEGTOQLGELINTECDTDG---- 115  
 103 -KCTRKSCRNPDPVNGMVHVI--KIDIFGSOJINSCNKKYRLIGS--SATCIIISGTVI 158  
 116 WTNDIPICEVVKC-LPYTAPENKIVSSAMEPDEYH-FGQAVRFVCSNGS-----YKIE 167  
 159 WDNETPICEIIPGCLPPTI-ANGDFISTS---REYFYSVYTYRNLGSGRKLFEIV 213

168 GDEEMHCS--DD--GFWSKEKPKC-VEISCKSPDVINGSPIS 204  
 214 GEPSTYCTSKNDQYGVMSGAPACIIPNKKMPVNGVILVS 255

RESULT 11  
 ID P87616 PRELIMINARY; PRT; 259 AA.  
 AC P87616;  
 DT 01-MAY-1997 (TREMBLrel. 03, Created)  
 DT 01-MAY-1997 (TREMBLrel. 03, last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, last annotation update)  
 DE 41KB FRAGMENT FROM LEFT END OF GENOME.  
 GN D17L OR C17L.  
 OS Cowpox virus (CEV).  
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
 OC Orthopoxvirus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-GRI-90;  
 RX MEDLINE; 97068532.  
 RA Safonov P.F., Petrov N.A., Riazankina O.I., Totmenin A.V.,  
 RA Shchelkunov S.N., Sandakhchiev L.S.;  
 RT "Genes of a circle of hosts for the cowpox virus."  
 RL Dokl. Akad. Nauk 349:829-833(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-GRI-90;  
 RX MEDLINE; 98229462.  
 RA Shchelkunov S.N., Safonov P.F., Totmenin A.V., Petrov N.A.,  
 RA Ryzankina O.I., Gutorov V.V., Kotwal G.J.;  
 RT "The genomic sequence analysis of the left and right species-specific  
 RT terminal region of a cowpox virus strain reveals unique sequences and  
 RT a cluster of intact ORFs for immunomodulatory and host range  
 RT proteins."  
 RL Virology 243:432-460(1998).  
 DR EMBL; X94355; CAA64102.1;  
 DR EMBL; Y11842; CAA23567.1;  
 DR HSSP; P10998; IVC.  
 DR INTERPRO; IPR000436;  
 DR PFM; PFO0084; sushi; 4.  
 DR SEQUENCE 259 AA; 28193 MW; 9D1AAEF68938859A CRC64;

Query Match 22.0%; Score 255.5; DB 12; Length 259;  
 Best Local Similarity 33.8%; Pred. No. 8.3e-18;  
 Matches 69; Conservative 23; Mismatches 93; Indels 19; Gaps 10;

3 CNEPPRRNTEILGWSMDQTYEGTQAIYKCRGYR--SLGNVIMVCRKGEWALNPLR 60  
 20 CCLPSPRITMKKGT-VDSHYNIGDIEVLCIPGYKQKMGPIYAKCTGTGTLFN--- 75  
 61 KQQRKPCGHPDTPPEGTFTLLTGAVNEFYGVKAVYTCNEGTOQLGELINTECDTDG---W 116  
 76 QCIKRCPCSPRIDNQLDIGG---VDFGSSITYSCNSGYHLIGESKSYCLGSLTSGSMV 132

117 TNDIPICEVVKCLPYTAPENKIVSSAMEPDEYHFGQAVRFVCSNGYKIEDEEMHCS 176  
 133 NEAPICESVKQSPSPISNGR--HNGYED--FYDGSVITYSCNSGSLIGNSGVLC- 187  
 177 DGFWSKEKPKC-VEISCKSPDVING 200  
 188 GGEWS-DPPTCOIVCPHPTISNG 210

RESULT 12  
 ID 090V2 PRELIMINARY; PRT; 559 AA.  
 AC 090V2;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, last annotation update)  
 DE HUMAN CRI mRNA FOR C3B/C4B RECEPTOR SECRETED FORM PRECURSOR (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 89010527.  
 RA Hourcade D., Miesner D.R., Atkinson J.P., Holers V.M.;  
 RT "Identification of an alternative polyadenylation site in the human  
 RT C3b/C4b receptor (complement receptor type 1) transcriptional unit and  
 RT prediction of a secreted form of complement receptor type 1."  
 RL J. Exp. Med. 168:1255-1270(1988).  
 DR EMBL; X14362; CAA32541.1;  
 DR INTERPRO; IPR000436;  
 DR INTERPRO; IPR000834;  
 DR PFM; PFO0084; sushi; 8.  
 DR PROSITE; PS00133; CARBOXYPEPT\_ZN\_2; UNKNOWN\_1.  
 KW Signal.  
 FT NON\_TER 1  
 FT SIGNAL 1  
 FT CHAIN <1 16 POTENTIAL.  
 FT CHAIN 17 559 POTENTIAL.  
 SQ SEQUENCE 559 AA; 61424 MW; DBPFE965CA179D75 CRC64;

Query Match 21.9%; Score 255; DB 4; Length 559;  
 Best Local Similarity 30.3%; Pred. No. 2.4e-17;  
 Matches 67; Conservative 39; Mismatches 83; Indels 32; Gaps 13;

3 CNE---LPPRRNTEILGWSMDQTYEGTQAIYKCRGYRSLGNVIMVCRKGEWALNPL 59  
 18 CNAPEWLFPARPNTLT-----DEFEPIGTINTYRCRGYSGRPSIICLNKSVWTSKAD- 72  
 60 RKCQRKPCGHPDTPPEGTFTLLTGAVNEFYGVKAVYTCNEGTOQLGELINTECDTDG---- 115  
 73 -KCTRKSCRNPDPVNGMVHVIK--IOFGSQIKYSCYKGYRLIGS--SATCIIISGTVI 128  
 116 WTNDIPICEVVKC-LPYTAPENKIVSSAMEPDEYHFGQAVRFVCSNGS-----YKIEG 168  
 129 WDNETPICEIIPGCLPPTI-TNGDFISTNRE---NFHGSVYTYRNLGSGRKFELV 184  
 169 DEEMHC--SD--GFWSKEKPKC-VEISCKSPDVINGSPIS 204  
 185 EPSTYCTSKNDQYGVMSGAPACIIPNKKCTPPNVNGVILVS 225

RESULT 13  
 ID 016745 PRELIMINARY; PRT; 2039 AA.  
 AC 016745;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, last annotation update)  
 DE COMPLEMENT RECEPTOR 1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN [1]

```
RP SEQUENCE FROM N.A.
RX MEDLINE; 94065175.
RA Vik D.P., Wong W.W.;
RT "Structure of the gene for the F allele of complement receptor type 1
RL J. Immunol. 151:6214-6224(1993).
RN [2]
RP SEQUENCE FROM N.A.
RA Vik D.P., Wong W.W.;
RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; L17418; AAB60694.1; JOINED.
DR EMBL; L17390; AAB60694.1; JOINED.
DR EMBL; L17391; AAB60694.1; JOINED.
DR EMBL; L17392; AAB60694.1; JOINED.
DR EMBL; L17393; AAB60694.1; JOINED.
DR EMBL; L17394; AAB60694.1; JOINED.
DR EMBL; L17395; AAB60694.1; JOINED.
DR EMBL; L17396; AAB60694.1; JOINED.
DR EMBL; L17397; AAB60694.1; JOINED.
DR EMBL; L17398; AAB60694.1; JOINED.
DR EMBL; L17399; AAB60694.1; JOINED.
DR EMBL; L17400; AAB60694.1; JOINED.
DR EMBL; L17401; AAB60694.1; JOINED.
DR EMBL; L17402; AAB60694.1; JOINED.
DR EMBL; L17403; AAB60694.1; JOINED.
DR EMBL; L17404; AAB60694.1; JOINED.
DR EMBL; L17405; AAB60694.1; JOINED.
DR EMBL; L17406; AAB60694.1; JOINED.
DR EMBL; L17407; AAB60694.1; JOINED.
DR EMBL; L17408; AAB60694.1; JOINED.
DR EMBL; L17409; AAB60694.1; JOINED.
DR EMBL; L17410; AAB60694.1; JOINED.
DR EMBL; L17411; AAB60694.1; JOINED.
DR EMBL; L17412; AAB60694.1; JOINED.
DR EMBL; L17413; AAB60694.1; JOINED.
DR EMBL; L17414; AAB60694.1; JOINED.
DR EMBL; L17415; AAB60694.1; JOINED.
DR EMBL; L17416; AAB60694.1; JOINED.
DR EMBL; L17417; AAB60694.1; JOINED.
DR EMBL; L17418; AAB60694.1; JOINED.
DR EMBL; L17419; AAB60694.1; JOINED.
DR EMBL; L17420; AAB60694.1; JOINED.
DR EMBL; L17421; AAB60694.1; JOINED.
DR EMBL; L17422; AAB60694.1; JOINED.
DR EMBL; L17423; AAB60694.1; JOINED.
DR HSSP; P08603; IHFI.
DR INTERPRO; IPR000436; -.
DR INTERPRO; IPR000834; -.
DR INTERPRO; IPR001424; -.
DR PFM; PF00084; sushi; 30.
DR PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN_1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_2.
SQ SEQUENCE 2039 AA; 223603 MW; B82FCB11C6B16635 CRC64;
```

```
Query Match 21.9%; Score 255; DB 4; Length 2039;
Best Local Similarity 30.3%; Pred. No. 1.1e-16;
Matches 67; Conservative 39; Mismatches 83; Indels 32; Gaps 13;

QY 3 CNE---LPKRNTELIGSWSQDYTPGSTQAIYKCRPGYRSIGVNYICRKGWVALNPL 59
DB 43 CNADEWLPFAFPTNLT-----DEFEFPICTLYNYECRPGYSCRPSILCLKSVWIGAD- 97
QY 60 RKCKRRPGHGDPTGPGFTLTGNGVFEYGVKAYTYTCGSGQLLGEINYYRCC---DTDG 115
DB 98 -RCRKRKGRNPPDPVNGVAVIKG--IOPGSOIKYSCRTGYRLIGS--SSAVCIISGDTVI 153
QY 116 WNDIPICEVVKC-LPYTAPENGRIVSAMPEDEYHFGQAVRFVNCSG-----YKIEG 168
DB 154 WNDETPICDRIPICGLPPTI-TNGDFISNRE---NFHYGSVYTYRCNPDGSGKRVFELVG 209
QY 169 DEEMHC--SDD--GFWSKEKPKC-VEISCKSPDVINGSPIS 204
DB 210 EPTSYCTSDNDQVIGWSPAPQCIIPNKCPTPPNVENGILWS 250

RESULT 14
ID 016744 PRELIMINARY; PRT: 2489 AA.
AC 016744;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE COMPLEMENT RECEPTOR 1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA Vik D.P., Wong W.W.;
RX MEDLINE; 94065175.
RA Vik D.P., Wong W.W.;
RT "Structure of the gene for the F allele of complement receptor type 1
RL J. Immunol. 151:6214-6224(1993).
RN [2]
RP SEQUENCE FROM N.A.
RA Vik D.P., Wong W.W.;
RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; L17418; AAB60695.1; JOINED.
DR EMBL; L17390; AAB60695.1; JOINED.
DR EMBL; L17391; AAB60695.1; JOINED.
DR EMBL; L17392; AAB60695.1; JOINED.
DR EMBL; L17393; AAB60695.1; JOINED.
DR EMBL; L17394; AAB60695.1; JOINED.
DR EMBL; L17395; AAB60695.1; JOINED.
DR EMBL; L17396; AAB60695.1; JOINED.
DR EMBL; L17397; AAB60695.1; JOINED.
DR EMBL; L17398; AAB60695.1; JOINED.
DR EMBL; L17399; AAB60695.1; JOINED.
DR EMBL; L17400; AAB60695.1; JOINED.
DR EMBL; L17401; AAB60695.1; JOINED.
DR EMBL; L17402; AAB60695.1; JOINED.
DR EMBL; L17403; AAB60695.1; JOINED.
DR EMBL; L17404; AAB60695.1; JOINED.
DR EMBL; L17405; AAB60695.1; JOINED.
DR EMBL; L17406; AAB60695.1; JOINED.
DR EMBL; L17407; AAB60695.1; JOINED.
DR EMBL; L17408; AAB60695.1; JOINED.
DR EMBL; L17409; AAB60695.1; JOINED.
DR EMBL; L17410; AAB60695.1; JOINED.
DR EMBL; L17411; AAB60695.1; JOINED.
DR EMBL; L17412; AAB60695.1; JOINED.
DR EMBL; L17413; AAB60695.1; JOINED.
DR EMBL; L17414; AAB60695.1; JOINED.
DR EMBL; L17415; AAB60695.1; JOINED.
DR EMBL; L17416; AAB60695.1; JOINED.
DR EMBL; L17417; AAB60695.1; JOINED.
DR EMBL; L17418; AAB60695.1; JOINED.
DR EMBL; L17419; AAB60695.1; JOINED.
DR EMBL; L17420; AAB60695.1; JOINED.
DR EMBL; L17421; AAB60695.1; JOINED.
DR EMBL; L17422; AAB60695.1; JOINED.
DR EMBL; L17423; AAB60695.1; JOINED.
DR EMBL; L17424; AAB60695.1; JOINED.
DR EMBL; L17425; AAB60695.1; JOINED.
DR EMBL; L17426; AAB60695.1; JOINED.
DR EMBL; L17427; AAB60695.1; JOINED.
DR EMBL; L17428; AAB60695.1; JOINED.
DR EMBL; L17429; AAB60695.1; JOINED.
DR EMBL; L17430; AAB60695.1; JOINED.
DR HSSP; P08603; IHFI.
DR INTERPRO; IPR000436; -.
DR INTERPRO; IPR000834; -.
DR INTERPRO; IPR001424; -.
DR PFM; PF00084; sushi; 37.
DR PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN_1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_3.
SQ SEQUENCE 2489 AA; 272846 MW; CE11B53F2B4FAF6 CRC64;
```





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2000, 16:43:42 ; Search time 49.92 Seconds  
(without alignments)  
181.518 Million cell updates/sec

Title: US-09-316-163-10

Perfect score: 1497

Sequence: 1 EDCNELPPRRNTEILTGSMS.....EKSCDNPYPNGDYSPLRIK 265

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_36:\*

1: /SIDSL/gcgdata/geneseq/geneseqp/AA1980.DAT:\*

2: /SIDSL/gcgdata/geneseq/geneseqp/AA1981.DAT:\*

3: /SIDSL/gcgdata/geneseq/geneseqp/AA1982.DAT:\*

4: /SIDSL/gcgdata/geneseq/geneseqp/AA1983.DAT:\*

5: /SIDSL/gcgdata/geneseq/geneseqp/AA1984.DAT:\*

6: /SIDSL/gcgdata/geneseq/geneseqp/AA1985.DAT:\*

7: /SIDSL/gcgdata/geneseq/geneseqp/AA1986.DAT:\*

8: /SIDSL/gcgdata/geneseq/geneseqp/AA1987.DAT:\*

9: /SIDSL/gcgdata/geneseq/geneseqp/AA1988.DAT:\*

10: /SIDSL/gcgdata/geneseq/geneseqp/AA1989.DAT:\*

11: /SIDSL/gcgdata/geneseq/geneseqp/AA1990.DAT:\*

12: /SIDSL/gcgdata/geneseq/geneseqp/AA1991.DAT:\*

13: /SIDSL/gcgdata/geneseq/geneseqp/AA1992.DAT:\*

14: /SIDSL/gcgdata/geneseq/geneseqp/AA1993.DAT:\*

15: /SIDSL/gcgdata/geneseq/geneseqp/AA1994.DAT:\*

16: /SIDSL/gcgdata/geneseq/geneseqp/AA1995.DAT:\*

17: /SIDSL/gcgdata/geneseq/geneseqp/AA1996.DAT:\*

18: /SIDSL/gcgdata/geneseq/geneseqp/AA1997.DAT:\*

19: /SIDSL/gcgdata/geneseq/geneseqp/AA1998.DAT:\*

20: /SIDSL/gcgdata/geneseq/geneseqp/AA1999.DAT:\*

21: /SIDSL/gcgdata/geneseq/geneseqp/AA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1055	70.5	240	18	Human partial Comp
2	961	64.2	216	18	Clone PRB9PR410 C
3	303	20.2	581	12	Human C4 binding P
4	301.5	20.1	263	20	Vaccinia complemen
5	301	20.1	496	20	Human CRI protein
6	300.5	20.1	543	13	CRI-4 (99H, 103E)
7	298.5	19.9	543	13	CRI-4 (52S, 53S, 5
8	298.5	19.9	543	13	CRI-4 (318-321 RNP
9	298.5	19.9	778	19	Amino acid sequenc
10	298.5	19.9	1930	19	Human complement r
11	298.5	19.9	2039	20	Human C3b/C4b rece
12	298.5	19.9	2317	10	CRI protein. Homo

13	296.5	19.8	543	13	R28560
14	295.5	19.7	543	13	R28570
15	294.5	19.7	543	13	R28550
16	294.5	19.7	543	13	R28553
17	294.5	19.7	543	13	R28565
18	294.5	19.7	543	13	R28568
19	294.5	19.7	543	13	R28569
20	294.5	19.7	543	13	R28571
21	294.5	19.7	2039	12	R1810
22	293.5	19.6	543	13	R28545
23	293.5	19.6	543	13	R28548
24	293	19.6	363	20	W12414
25	293	19.6	363	18	X30918
26	292.5	19.5	543	13	R28555
27	292.5	19.5	543	13	R28558
28	292.5	19.5	543	13	R28553
29	291.5	19.5	543	13	R28549
30	291.5	19.5	543	13	R28551
31	291.5	19.5	543	13	R28562
32	291.5	19.5	543	13	R28566
33	290.5	19.4	254	15	R47154
34	290.5	19.4	254	15	R47155
35	289.5	19.3	263	10	P92003
36	289.5	19.3	543	13	R28556
37	289.5	19.3	543	13	R28554
38	288.5	19.3	543	13	R28552
39	288.5	19.3	543	13	R28561
40	287.5	19.2	543	13	R28554
41	286.5	19.1	543	13	R28544
42	286.5	19.1	543	13	R28546
43	285.5	19.1	543	13	R28543
44	285.5	19.1	543	13	R28559
45	284.5	19.0	263	20	T29858

#### ALIGNMENTS

RESULT 1	
W39154	W39154 standard; Protein: 240 AA.
XX	
AC	W39154:
XX	
DE	Human partial Complement factor H protein fragment 1.
XX	
XX	Complement factor H; tumour associated antigen; renal cancer;
KW	urogenital cancer; medicament; modulator.
XX	
OS	Homo sapiens.
XX	
PN	WO9738136-A1.
XX	
PD	16-OCT-1997.
XX	
PF	09-APR-1997; 97WO-US05710.
XX	
PR	06-MAR-1997; 97US-0812481.
PR	09-APR-1996; 96US-0015083.
PR	09-APR-1996; 96US-0630048.
PR	06-MAR-1997; 97US-0038614.
XX	
PA	(BARD-) BARD DIAGNOSTIC SCI INC.
XX	
PI	Enfield DL, Hass GM, Kinders RJ;
XX	
DR	WPI, 1997-512742/47.
XX	
PT	N-PDB: V02790.
PT	Treating or screening for cancer, e.g. renal or urogenital cancer -
	by modulating or detecting tumour associated human complement Factor

PT H related antigen, or nucleic acid encoding it  
 XX  
 PS Example 6B; Fig 6B; 104pp: English.  
 CC This partial protein sequence represents a region of the human  
 CC tumour-associated complement factor H (CFH). This sequence is used  
 CC in the identification of complement factor H related proteins and  
 CC antigens isolated from clone PRB9FH410 (see W39155). The detection of  
 CC such proteins and a CFH antigens can be used in screening or for the  
 CC treatment of renal or urogenital cancer, e.g. bladder, cervical or  
 CC prostate cancer. Agents that may modulate this antigen could be used in  
 CC the manufacture of a medicament for the treatment of a tumour cell.  
 CC  
 SO Sequence 240 AA;  
 Query Match 70.5%; Score 1055; DB 18; Length 240;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-81;  
 Matches 188; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 78 FTLTGVNFEYGVKAVYTCNEGYLLGEINRREDTGTGWIINDIPICEVVKCLPYTAPENG 137  
 Db 1 flltgvnfeiygvkavylcnegyllgeinrredtdgtwtndipicevkvclpytapeng 60  
 QY 138 KIYSSAMEPDRHYHFGQAVRFVCSGKRIEDDEMHGSDGDFWSEKPKCYEISCKSPDV 197  
 Db 61 kiysamepdrhyhfgqavrfvcsngykriegdeemhcsdgdgfwsekpkyeiscskspdv 120  
 QY 198 INSPISQKTIYKENERFQYCKNMGYSEKRGAVCTESGWRPLPSCSEKSCDNPYIPNG 257  
 Db 121 insplsqkilykenerfyqckmgysekgavctesgwrplpsceekscdnpyping 180  
 QY 258 DYSPLRIK 265  
 Db 181 dysplrik 188  
 RESULT 2  
 ID W39155 standard; Protein; 216 AA.  
 AC W39155;  
 DT 27-APR-1998 (first entry)  
 DE Clone PRB9FH410 CFH related protein fragment.  
 KW Complement factor H; tumour associated antigen; renal cancer;  
 KW urogenital cancer; medicament; modulator.  
 OS Synthetic.  
 PN WO9738136-A1.  
 PD 16-OCT-1997.  
 PF 09-APR-1997; 97WO-US05710.  
 PR 06-MAR-1997; 97US-0812481.  
 PR 09-APR-1996; 96US-0015083.  
 PR 09-APR-1996; 96US-0630048.  
 PR 06-MAR-1997; 97US-0038614.  
 PA (BARD-) BARD DIAGNOSTIC SCI INC.  
 PI Enfield DL, Hass GM, Kinders RJ;  
 DR WPI: 1997-512742/47.  
 DR N-PSDB: V02791.  
 PT Treating or screening for cancer, e.g. renal or urogenital cancer -  
 PT by modulating or detecting tumour associated human Complement Factor  
 PT H related antigen, or nucleic acid encoding it

XX  
 PS Example 6B; Fig 6B; 104pp: English.  
 CC This partial protein is found in clone PRB9FH410 and represents a  
 CC complement factor H related protein with homology to a region of the  
 CC human tumour-associated complement factor H (CFH). The detection of this  
 CC protein and a CFH antigen can be used in screening or for the treatment  
 CC of renal or urogenital cancer, e.g. bladder, cervical or prostate cancer.  
 CC Agents that may modulate this antigen could be used in the manufacture of  
 CC a medicament for the treatment of a tumour cell.  
 CC  
 SO Sequence 216 AA;  
 Query Match 64.2%; Score 961; DB 18; Length 216;  
 Best Local Similarity 99.4%; Pred. No. 2.2e-73;  
 Matches 170; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 95 TCNNGYQLLGINRREDTGTGWIINDIPICEVVKCLPYTAPENGKIYSSAMEPDRHYHFGQ 154  
 Db 1 tcnngyqllgelnrredtdgtwtndipicevkvclpytapengkiysamepdrhyhfgq 60  
 QY 155 AVRFVCSGKRIEDDEMHGSDGDFWSEKPKCYEISCKSPDVINGSPISQKIYKENER 214  
 Db 61 avrfvcsngykriegdeemhcsdgdgfwsekpkyeiscskspdvinsplsqkilykener 120  
 QY 215 FQYCKNMGYSEKRGDAVCTESGWRPLPSCSEKSCDNPYIPNGDYSPLRIK 265  
 Db 121 fqyckmgysekgavctesgwrplpsceekscdnpypingdysplrik 171  
 RESULT 3  
 ID R13490 standard; Protein; 581 AA.  
 AC R13490;  
 DT 30-OCT-1991 (first entry)  
 DE Human C4 binding protein.  
 KW C4bp; monomer; complement protein; pJOD.C4bp.3; SCR;  
 KW short consensus repeat.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FH Peptide 1..32  
 FT /label= signal\_peptide  
 FT Protein 33..581  
 FT /label= C4bp  
 FT Region 33..93  
 FT /label= SCR8  
 FT Region 94..155  
 FT /label= SCR7  
 FT Region 156..219  
 FT /label= SCR6  
 FT Region 220..279  
 FT /label= SCR5  
 FT Region 280..345  
 FT /label= SCR4  
 FT Region 346..406  
 FT /label= SCR3  
 FT Region 407..464  
 FT /label= SCR2  
 FT Region 465..523  
 FT /label= SCR1  
 FT Domain 524..581  
 FT /label= C4bp-core  
 FT /note= "responsible for multimer assembly"  
 FT Disulfide-bond 34..80  
 FT /note= "intradomain"  
 FT Disulfide-bond 65..92

[illegible]

Query Match	20.1%	Score 301.5	DB 20	Length 263
Best Local Similarity	33.3%	Pred. No. 6,2e-18		
Matches 78	Conservative 29	Mismatches 106	Indels 21	Gaps 12
20	SPQTYPEGQATLYKCRGR--SLGNVIMWCKGGEWVALPLRKQCKRRCGHGDPFPGT	77		
DB	40	ananyiggtietleylclpgryrkqkmpiyakctgtgtwlf--qclktcpsprldnsg	96	
OY	78	FULTGNGVEYEGKAVAYTQNEGQYLGEL-NVRECDTDC---WTNDIPICEVVKCLPVYA	133	
DB	97	ldlsg---vdfigsltsycnsngyhligesksyglsgelsgtsmwnpaplcseckcspps	153	
OY	134	PENGKIVSSAMEPDRYHFGQAVRFVNCNGYKITEGDEEMHCSDPGWMSKEXKRCVETISK	193	
DB	154	lsnrg--hngyed--fyldgsvvlyscnsysllngsvlcs--ggews--dptclqivkcp	207	
OY	194	SPDVINGSPISO-KITIKENERPOFYCKNNGYEVSEBGDVCETSG--WRP-LPSC	244	
DB	208	hptlsgnylssgfkrsydnvdftckrygkrlsgssstcpgnkwkpelkpc	261	

RESULT 5  
 Y55752 standard: Protein: 496 AA.  
 AC Y55752;  
 DT 22-FEB-2000 (first entry)  
 DE Human CRI protein LHR-A SCR fragment.  
 XX  
 XX C3B/C4B receptor; CRI protein; cell-surface protein; erythrocyte; human;  
 KW complement regulatory activity; complement pathway enzyme; tissue damage;  
 KW reperfusion injury; Arthus reaction; myocardial infarct; inflammation;  
 KW heart condition; autoimmune disorder; long homologous repeat; LHR; SCR;  
 KW short consensus repeat.  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 XX US961481-A.  
 PN  
 PD 09-NOV-1999.  
 XX  
 XX 06-JUN-1995; 95US-0470652.  
 PF  
 PR 03-APR-1989; 89US-0332865.  
 PR 06-DEC-1974; 74US-0350238.  
 PR 24-FEB-1993; 93US-0026134.  
 PR 01-APR-1988; 88US-0176532.  
 XX  
 XX (UYJO ) UNIV JOHNS HOPKINS.  
 PA (BGHM ) BRIGHAM & WOMENS HOSPITAL.  
 PA (AVAN-) AVANT IMMUNOTHERAPEUTICS INC.  
 XX  
 PI Concino MF, Wong MW, Makrides SC, Klickstein LB, Fearon DT, IP SH;  
 PI Marsh HC, Carson GR;  
 DR WPI: 1999-633357/54.  
 XX  
 PT A human C3B/C4B receptor (CRI) protein having antinflammatory and  
 PT cardiant activity -  
 XX  
 PS Disclosure: Fig 10; 87pp: English.  
 XX  
 XX The invention relates to a human C3B/C4B receptor (CRI) protein. The CRI  
 CC protein or fragment is expressed as a cell-surface protein on the surface  
 CC of a non-human cell and exhibits a complement regulatory activity of full  
 CC length human CRI as expressed on erythrocytes. The CRI function in vivo  
 CC may be mediated through the inhibition of complement pathway enzymes. The  
 CC soluble CRI protein exhibits a complement regulatory activity, and this  
 CC may be used to prevent reperfusion injury, inhibit Arthus reaction, and  
 CC neutrophil mediated tissue damage, and reduce myocardial infarct size,  
 CC and inflammation. The CRI protein and its fragments can also be used in  
 CC the treatment of conditions which involve unwanted complement activity,  
 CC e.g. shock lung, tissue damage due to burn, or ischemic heart conditions,  
 CC and autoimmune disorders. CRI proteins, analogues, derivatives, and anti-  
 CC CRI antibodies are used in assays, and diagnostics. The present sequence  
 CC represents the short consensus repeat (SCR) fragments of human CRI  
 CC protein long homologous repeat (LHR)-A sequence.  
 XX  
 SQ Sequence 496 AA:

Query Match 20.1%; Score 301; DB 20; Length 496;  
 Best Local Similarity 27.5%; Pred. No. 1.5e-17;  
 Matches 88; Conservative 48; Mismatches 110; Indels 74; Gaps 19;

QY 3 CNE---LPPRNTETLTGSGWSQDTPEGTQAIYKCRPGYGRSGNIMWCRKEB-WVALNP 58  
 DB 48 cnapewlfarptult---defefpigtlylncycrpgysgrpfslcclksnswtgakd 103  
 QY 59 LKRCOKRPGCGHGDTPFGFTLTGGNVPEYGVKAVYTCNEGYQLGEIMYRRC---DND 114  
 DB 104 --rcrkscrpdpvngmwhvikg--igtgsqlyksckkyrlilgs--ssatclisgdtlv 158

QY 115 GWINDIPICEYVWC-LPVLAPENGKIVSSAMEPDREYHFGQAVRFVNSG-----YKIE 167  
 DB 159 iwdnetpdcdrpcglpplcl-tngdfistnre---nfhygsvnlrcmpnsggkvtel 214  
 QY 168 GDEEMHC--SDD--GFWSEKPKC-VEISCKSPDYVINGSPIS-QKTIYKENERQYCNM 221  
 DB 215 gepaiyctsnddqvgylwspgpcglpnkcclpnhvengllvsdnrslfslneveftcqp 274  
 QY 222 GYEYSEKGDVACTE-SCWRP-LPSCDE-----KSCDN 251  
 DB 275 gfvmkgyrrvkcqqlnkwepeipscsvvcqpppvlnaertqrkdnfsgqevfyfscp 334  
 QY 252 PY-----IPNGDYSP 261  
 DB 335 gydlrgaasmrcipgqdwsp 354

RESULT 6  
 R28557  
 ID R28557 standard: peptide: 543 AA.  
 XX  
 AC R28557;  
 XX  
 DT 19-MAR-1993 (first entry)  
 XX  
 DE CRI-4 (99H, 103E) analogue.  
 XX  
 KW short consensus repeat; regulator of complement activation;  
 KW C3b binding; C4b binding; human complement type 1 receptor.  
 XX  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FH Region 1..60  
 FT /label= SCR-1  
 FT Region 61..122  
 FT /label= SCR-2  
 FT Region 451..510  
 FT /label= SCR-8  
 FT Region 511..543  
 FT /label= SCR-9  
 FT /note= "TRUNCATED"  
 FT Misc-difference 99  
 FT /note= "Ser substituted by His (SCR-9)"  
 FT FT Misc-difference 103  
 FT /note= "Thr substituted by Glu (SCR-9)"  
 PN EP512733-A.  
 XX  
 PD 11-NOV-1992.  
 XX  
 PF 28-APR-1992; 92EP-0303826.  
 XX  
 PR 03-MAY-1991; 91US-0695514.  
 XX  
 PA (UNTW ) UNIV WASHINGTON.  
 XX  
 PI Atkinson JP, Hourcade D, Krych M;  
 DR WPI: 1992-375009/46.  
 XX  
 PT Complement activity regulator protein analogues - useful for  
 PT treating autoimmune diseases, to suppress transplant rejection,  
 PT for diagnosis etc.  
 XX  
 PS Claim 11; Fig 2 and R11810; 23pp: English.  
 XX  
 CC The cDNA clone designated CRI-4 was described in J. Exp. Med. (1988)  
 CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
 CC SCRs of CRI. The invention concerns analogues of "regulator of  
 CC complement activation" proteins or truncated, hybrid or recombinant  
 CC forms of them. CRI-4 is a preferred truncated form and a number of



XX Short consensus repeat; regulator of complement activation;  
 KW C3b binding; C4b binding; human complement type 1 receptor.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 1..60  
 FT /label= SCR-1  
 FT Region 61..122  
 FT /label= SCR-2  
 FT Region 451..510  
 FT /label= SCR-8  
 FT Region 511..543  
 FT /label= SCR-9  
 FT /note= "TRUNCATED"  
 FT Misc-difference 318  
 FT /note= "Asp substituted by Arg"  
 FT Misc-difference 319  
 FT /note= "Asp substituted by Asn"  
 FT Misc-difference 320  
 FT /note= "Phe substituted by Pro"  
 FT Misc-difference 321  
 FT /note= "Met substituted by Pro"  
 XX  
 XX EP512733-A.  
 XX  
 XX 11-NOV-1992.  
 XX  
 XX 28-APR-1992; 92EP-0303826.  
 XX  
 XX 03-MAY-1991; 91US-0695514.  
 XX  
 XX (UNITM ) UNIV WASHINGTON.  
 XX  
 XX Atkinson JP, Hourcade D, Krych M;  
 PI WPI: 1992-375009/46.  
 XX  
 XX Complement activity regulator protein analogues - useful for  
 PT treating auto-immune diseases, to suppress transplant rejection,  
 PT for diagnosis etc.  
 XX  
 XX Example 8; Page 18 and R11810; 23pp; English.  
 XX  
 XX The cDNA clone designated CRI-4 was described in J. Exp. Med. (1988)  
 CC 168:1235-1270. It encodes the first 8 and a half amino terminal  
 CC SCRs of CRI. The invention concerns analogues of "regulator of  
 CC complement activation" proteins or truncated, hybrid or recombinant  
 CC forms of them. CRI-4 is a preferred truncated form and a number of  
 CC specified substitution variants of it are disclosed in which certain  
 CC positions in SCR-5-6 are substituted by amino acids from  
 CC the corresponding positions in SCRs which are involved in C3b- and  
 CC C4b-binding. The substitution variant given here has increased C4b-  
 CC binding. The specification does not contain the CRI-4 sequence;  
 CC the sequence given here was constructed from the full-length CRI  
 CC amino acid sequence having GENESEQ accession number R11810 and  
 CC descriptions in the disclosure.  
 XX  
 XX Sequence 543 AA:  
 SQ  
 Query Match 19.9%; Score 298.5; DB 13; Length 543;  
 Best Local Similarity 26.5%; Pred. No. 2.7e-17;  
 Matches 88; Conservative 44; Mismatches 105; Indels 95; Gaps 18;  
 Oy 3 CNE---LPRRMTETLTGSMSTQTYPEGTQATYKCRPGYRSLGNVIMVCRKGEVALNDL 59  
 Db 2 cnapewlpfarpfntlt---defefpigtlylnyecpysgqpfisiclknswtgakt- 56  
 Oy 60 KRCQKRPCHGPDPTPGTFTTLTGCVNFEYGVAVYTCNNGYQLGELINREC---DTDG 115  
 Db 57 -rcrrksctnppdpvngmhwikg--lqfsgqskysctkyrligs-saatclsgdtvl 112

Oy 116 WTNDIPICEVVK-LPVTADENGKIVSSAMEPPDEHFHGOAVRVCNSG-----YKIEG 168  
 Db 113 wdnetpicdrtpoglpplrt-tngdfistne---nfhgsvvlyrcnpgsgqtkvfeivg 168  
 Oy 169 DEEMHC--SDD--GFMSKEPKC-VEISCKSPDVINCSPIS-QRIYKEN----- 212  
 Db 169 epslyctsnddqygiwspapqclipnktcpnvengllvsdnrlsflsnevefrcpqv 228  
 Oy 213 -----ERFO-----YKCNMG 222  
 Db 229 fvmkgprtrvkcqalnkwpelpscsvrcqppdvllhaertqrdkdnfsgqevfyscpq 288  
 Oy 223 YEXSERGDAVCTESG-WRP-LPSCSEKSCDNP 252  
 Db 289 ydlrgaasmrcitpvgdksppaipcvcvscnrt 320  
 RESULT 9  
 ID W73147  
 XX W73147 standard; protein; 778 AA.  
 XX  
 AC W73147;  
 XX  
 DT 29-JAN-1999 (first entry)  
 XX  
 XX Amino acid sequence of the soluble complement receptor 1 (SCR1).  
 XX  
 XX Human; soluble complement receptor 1; SCR1; T-cell; B-cell;  
 KW mediated immune response; inhibition; tissue rejection; gene therapy;  
 KW dystrophin; inflammatory response; interferon-gamma secretory response;  
 KW autoimmune response; neurological response; Alzheimer's disease;  
 KW Parkinson's disease; multiple sclerosis; systemic lupus erythematosus;  
 KW rheumatoid arthritis; myasthenia gravis; epidermis bullosa;  
 KW Hashimoto's disease.  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 XX WO9845430-A1.  
 XX  
 XX 15-OCT-1998.  
 XX  
 XX 06-APR-1998; 98WO-GB01012.  
 XX  
 XX 05-APR-1997; 97GB-0006950.  
 XX  
 XX (ANNE/) ANNENKOV A.  
 PA (CHER/) CHERNAJOVSKY Y.  
 XX  
 XX Annenkov A, Chernajovsky Y;  
 PI WPI: 1998-568350/48.  
 XX  
 XX Fragment of soluble human complement receptor 1 - useful for  
 PT treating T-cell or B-cell mediated immune responses e.g.  
 PT inflammatory responses such as rheumatoid arthritis  
 XX  
 XX Disclosure; Fig 1; 54pp; English.  
 XX  
 XX This is an amino acid sequence of the human soluble complement  
 CC receptor 1 (SCR1), useful in the treatment of T-cell or B-cell  
 CC mediated immune responses. It is used to inhibit a T-cell or  
 CC B-cell-mediated immune response to prevent immune response-mediated  
 CC tissue rejection and destruction or clearance or inactivation of an  
 CC expressed protein especially from cells that have been treated by gene  
 CC therapy to express the protein, e.g. dystrophin. The protein can also  
 CC be used to inhibit a T-cell or B-cell-mediated inflammatory response,  
 CC an interferon-gamma secretory response, autoimmune response or  
 CC neurological response, e.g. Alzheimer's or Parkinson's disease or  
 CC multiple sclerosis. Also the protein can be used to treat systemic  
 CC lupus erythematosus, rheumatoid arthritis, myasthenia gravis,  
 CC epidermis bullosa or Hashimoto's disease.  
 XX



```

PR 06-DEC-1974: 74US-0350238.
PR 24-FEB-1993: 93US-0026134.
PR 01-APR-1988: 88US-0176532.
XX
XX (UYJO ) UNIV JOHNS HOPKINS.
XX PA (BGHM ) BRIGHAM & WOMEN'S HOSPITAL.
XX PA (AVAN-) AVANT IMMUNOTHERAPEUTICS INC.
XX
PI Concino MF, Wong WM, Makrides SC, Klickstein LB, Fearon DT, Ip SH;
PI Marsh HC, Carson GR;
XX
XX MPI: 1999-633357/54.
XX DR N-PSDB; 238150.
XX
XX A human C3B/C4B receptor (CRI) protein having antiinflammatory and
XX PT cardiant activity -
XX
XX Disclosure: Fig 1A-P; 87pp; English.
XX
XX The invention relates to a human C3B/C4B receptor (CRI) protein. The CRI
XX CC protein or fragment is expressed as a cell-surface protein on the surface
XX CC of a non-human cell and exhibits a complement regulatory activity of full
XX CC -length human CRI as expressed on erythrocytes. The CRI function in vivo
XX CC may be mediated through the inhibition of complement pathway enzymes. The
XX CC soluble CRI protein exhibits a complement regulatory activity, and this
XX CC may be used to prevent reperfusion injury, inhibit Arthus reaction, and
XX CC neutrophil mediated tissue damage, and reduce myocardial infarct size,
XX CC and inflammation. The CRI protein and its fragments can also be used in
XX CC the treatment of conditions which involve unwanted complement activity,
XX CC e.g. shock lung, tissue damage due to burn, or ischemic heart conditions,
XX CC and autoimmune disorders. CRI proteins, analogues, derivatives, and anti
XX CC -CRI antibodies are used in assays, and diagnostics. The present sequence
XX CC represents the human CRI protein.
XX
XX Sequence 2039 AA:
SQ
Query Match 19.9%; Score 298.5; DB 20; Length 2039;
Best Local Similarity 27.0%; Pred. No. 1.4e-16;
Matches 86; Conservative 49; Mismatches 111; Indels 73; Gaps 18;
QY 3 CNE---LPPRRNTTEILGSMDSQTYPEGTOAIYKCRPGYSLGNVIMVCKRGWALNPL 59
DB || || || || || || || || || || || || || || || || || || || ||
43 cnapewlpfarplntl----defefpdytlyncrpyysgrpfslclknswtgakd- 97
QY 60 RKQKRPCCGHPGTPGTFTLTGNGNFEYGVKAVYTCNREGYOLGELINREC-----DTDG 115
DB || || || || || || || || || || || || || || || || || || || ||
98 -rrrrscnppdpvngmwhvlykg--lqfsgqikysctkyrllgs--saatcllsgdtvl 153
QY 116 WTNDIPICEVVKC-LPVTAPENCKIIVSAMEPDREYHFGQAVRFVNCNG-----YKIEG 168
DB || || || || || || || || || || || || || || || || || || || ||
154 wdnefpdclripplf-tngdfistnre--nhfysvlyrcnpsggrkvfelvg 209
QY 169 DEEMHC--SDD--GFWSKKEPKC-VEISCKSPVYINGSPIS-OKIYKENERQYVCNMG 222
DB || || || || || || || || || || || || || || || || || || || ||
210 epsilyctsnddvgylwsgapqcllpnkctpvengllvsdmsrlfsjnevevfrcpg 269
QY 223 YEYSESGDAVCTE-SGWRP-LPSCDE-----KSCDNP 252
DB || || || || || || || || || || || || || || || || || || || ||
270 fymkgsprirvkqalnkwepeelpscsrvcqppdvhaertgrdkdnfsgqevfyscepg 329
QY 253 Y-----IPNGDYSP 261
DB || || || || || || || || || || || || || || || || || || || ||
330 ydlrgaasmrctpgqdwsp 348
DB
RESULT 12
P92219
ID P92219 standard; protein; 2317 AA.
XX
XX P92219;
XX
XX 22-FEB-1990 (first entry)

```

```

XX
XX CRI protein.
DE Complement; cofactor.
XX
XX Homo sapiens (human).
XX
XX Key location/Qualifiers
XX FT Peptide 10..50
XX FT /label= signal_peptide
XX
XX WO8909220-A.
XX
XX 05-OCT-1989.
XX
XX 31-MAR-1989; 89WO-US01358.
XX
XX 01-APR-1988; 88US-0176532.
XX
XX (TCEL ) T CELL SCIENCES INC.
XX PA (UYJO ) THE JOHNS HOPKINS UNIVERSITY.
XX PA (BRIG ) THE BRIGHAM AND WOMEN'S HOSPITAL.
XX
XX Fearon DT, Klickstein LB, Wong W, Carson G, Concino MF, Makrides SC;
XX DR MPI; 1989-309498/42.
XX DR N-PSDB; N91477.
XX
XX New nucleic acid sequences encoding new CRI protein - and its fragment,
XX PT for diagnosis and control of complement-related immune defects,
XX PT inflammation, myocardial infarct, etc
XX
XX Claim 1; Fig 1; 191pp; English.
XX
XX This is full-length CRI protein, and shortened forms are new, lacking
XX CC the transmembrane region. The proteins and fragments bind C3b and/or
XX CC C4b, have cofactor activity and inhibit C3 and C5 convertase activity.
XX CC In the sequence, x=untranslated region. This has 7 short consensus
XX CC repeats (CR) of 4 long direct homologous repeats (LHR) (see fig. 10 in
XX CC patent). Deletion mutants, eg LHR-A for binding of C4b, and LHR-C
XX CC for C3b binding, be made, and LHR-B and -C can detect CR2 sequences.
XX CC They are useful in diagnosing and treating immune disorders, and prevent
XX CC perfusion injury.
XX
XX Sequence 2317 AA:
SQ
Query Match 19.9%; Score 298.5; DB 10; Length 2317;
Best Local Similarity 27.0%; Pred. No. 1.6e-16;
Matches 86; Conservative 49; Mismatches 111; Indels 73; Gaps 18;
QY 3 CNE---LPPRRNTTEILGSMDSQTYPEGTOAIYKCRPGYSLGNVIMVCKRGWALNPL 59
DB || || || || || || || || || || || || || || || || || || || ||
52 cnapewlpfarplntl----defefpdytlyncrpyysgrpfslclknswtgakd- 106
QY 60 RKQKRPCCGHPGTPGTFTLTGNGNFEYGVKAVYTCNREGYOLGELINREC-----DTDG 115
DB || || || || || || || || || || || || || || || || || || || ||
107 -rrrrscnppdpvngmwhvlykg--lqfsgqikysctkyrllgs--saatcllsgdtvl 162
QY 116 WTNDIPICEVVKC-LPVTAPENCKIIVSAMEPDREYHFGQAVRFVNCNG-----YKIEG 168
DB || || || || || || || || || || || || || || || || || || || ||
163 wdnefpdclripplf-tngdfistnre--nhfysvlyrcnpsggrkvfelvg 218
QY 169 DEEMHC--SDD--GFWSKKEPKC-VEISCKSPVYINGSPIS-OKIYKENERQYVCNMG 222
DB || || || || || || || || || || || || || || || || || || || ||
219 epsilyctsnddvgylwsgapqcllpnkctpvengllvsdmsrlfsjnevevfrcpg 278
QY 223 YEYSESGDAVCTE-SGWRP-LPSCDE-----KSCDNP 252
DB || || || || || || || || || || || || || || || || || || || ||
279 fymkgsprirvkqalnkwepeelpscsrvcqppdvhaertgrdkdnfsgqevfyscepg 338
QY 253 Y-----IPNGDYSP 261
DB || || || || || || || || || || || || || || || || || || || ||

```



168:1255-1270. It encodes the first 8 and a half amino terminal SCRs of CRI. The invention concerns analogues of "regulator of complement activation" proteins or truncated, hybrid or recombinant forms of them. CRI-4 is a preferred truncated form and a number of specified substitution variants of it are disclosed in which certain positions in SCR-5-6 are substituted by amino acids from the corresponding positions in SCRs which are involved in C3b- and C4b-binding. The substitution variant given here has increased C3b-binding. The specification does not contain the CRI-4 sequence; the sequence given here was constructed from the full-length CRI amino acid sequence having GENESEQ accession number R11810 and descriptions in the disclosure.

Sequence 543 AA:

Query Match 19.7%; Score 295.5; DB 13; Length 543;  
Best Local Similarity 26.0%; Pred. No. 4.8e-17;  
Matches 86; Conservative 44; Mismatches 106; Indels 95; Gaps 17;

QY 3 CNE---LPPRRMTETITGSMSPDOTYPEGTOATYKCRPGYRSLGNVIMCRKGEWALNPL 59  
DB 2 cnapewlfarptnlt-----defefipgitylneyecrpgysgrpfslclknswtgcakd- 56  
QY 60 KRCQKRPCGHEGDPFGFTLTGNAVFEYGVKAVYTCNEGQVLGEINRYREC---DTDG 115  
DB 57 -rrrkscknppdpvngmhwikg--lqfsgqikyscktyrllgs-satclisgdlvi 112  
QY 116 WNDPIPCVVKC-LPVYAPENGKIVSSAMEPDREYHFGQAVRFVCSNG-----YKIEG 168  
DB 113 wneplcdrlpcqilppl-tngdfisture---nfhygsvvlyrcnpgsggrkvfelvg 168  
QY 169 DEEMHC--SD--GFWSKEKPKC-VEISCKSPDYINGSPIS-OKIYKEN----- 205  
DB 169 epslyctsnddqvlgwspapqcllpnkctprnvengllvsdnrlfslnevefrcqpv 228  
QY 206 -----KTIYKENRFQ-----YKCNMG 222  
DB 229 fvmkgprrvkcqalnkwepelpscsrvcqpppdlvhaertqrdknfsgqvfyfscpg 288  
QY 223 YEYSEKGDVCTESG-WRP-LPSCSEKSCDN 251  
DB 289 ydlrgaasmrctpgdwsapaapltcevkscdd 319

RESULT 15  
R28550 ID R28550 standard; peptide: 543 AA.  
XX R28550;

DT 19-MAR-1993 (first entry)  
XX CRI-4 (64K) analogue.

DE Short consensus repeat; regulator of complement activation;  
XX C3b binding; C4b binding; human complement type 1 receptor.  
KM C3b binding; C4b binding; human complement type 1 receptor.  
XX Homo sapiens.

OS Homo sapiens.  
XX Key Location/Qualifiers  
FH Region 1..60  
FT /label= SCR-1  
FT Region 61..122  
FT /label= SCR-2  
FT Region 451..510  
FT /label= SCR-8  
FT Region 511..543  
FT /label= SCR-9  
FT /note= "TRUNCATED"  
FT Misc-difference 64  
XX /note= "Arg substituted by Lys (SCR-9)"

PN EP512733-A.  
XX 11-NOV-1992.  
PD 28-APR-1992; 92EP-0303826.  
PF 03-MAY-1991; 91US-0695514.  
PR (UNIW ) UNIV WASHINGTON.  
XX Atkinson JP, Hourcade D, Krych M;  
XX WPI; 1992-375009/46.  
XX Complement actively regulator protein analogues - useful for  
PT treating auto-immune diseases, to suppress transplant rejection,  
PT for diagnosis etc.  
PS Claim 11; Fig 2 and R11810; 23pp; English.

The CDNA clone designated CRI-4 was described in J Exp Med.(1988)  
CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCRs of CRI. The invention concerns analogues of "regulator of  
CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed in which certain  
CC positions in SCR-2 which have been identified as important for the  
CC degree of C3b- and C4b-binding are substituted by amino acids from  
CC the corresponding positions in SCR-9. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENESEQ  
CC accession number R11810 and descriptions in the disclosure.

Sequence 543 AA:

Query Match 19.7%; Score 294.5; DB 13; Length 543;  
Best Local Similarity 26.3%; Pred. No. 5.8e-17;  
Matches 87; Conservative 45; Mismatches 104; Indels 95; Gaps 18;

QY 3 CNE---LPPRRMTETITGSMSPDOTYPEGTOATYKCRPGYRSLGNVIMCRKGEWALNPL 59  
DB 2 cnapewlfarptnlt-----defefipgitylneyecrpgysgrpfslclknswtgcakd- 56  
QY 60 KRCQKRPCGHEGDPFGFTLTGNAVFEYGVKAVYTCNEGQVLGEINRYREC---DTDG 115  
DB 57 -rrrkscknppdpvngmhwikg--lqfsgqikyscktyrllgs-satclisgdlvi 112  
QY 116 WNDPIPCVVKC-LPVYAPENGKIVSSAMEPDREYHFGQAVRFVCSNG-----YKIEG 168  
DB 113 wneplcdrlpcqilppl-tngdfisture---nfhygsvvlyrcnpgsggrkvfelvg 168  
QY 169 DEEMHC--SD--GFWSKEKPKC-VEISCKSPDYINGSPIS-OKIYKEN----- 212  
DB 169 epslyctsnddqvlgwspapqcllpnkctprnvengllvsdnrlfslnevefrcqpv 228  
QY 213 -----ERFQ-----YKCNMG 222  
DB 229 fvmkgprrvkcqalnkwepelpscsrvcqpppdlvhaertqrdknfsgqvfyfscpg 288  
QY 223 YEYSEKGDVCTESG-WRP-LPSCSEKSCDN 251  
DB 289 ydlrgaasmrctpgdwsapaapltcevkscdd 319

Search completed: November 21, 2000, 16:43:45  
Job time: 284 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2000, 16:49:21 ; Search time 27.94 Seconds  
(without alignments)  
376.194 Million cell updates/sec

Title: US-09-316-163-11

Perfect score: 1876  
Sequence: 1 EDCNELPPRRNTELTGSGWS.....PDIKHGGLYHNNRRPPFPV 329

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 87993 seqs, 31947931 residues

Total number of hits satisfying chosen parameters: 87993

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1876	100.0	1231	1 CFAH_HUMAN	P08603 homo sapien
2	1261	67.2	1234	1 CFAH_MOUSE	P06909 mus musculu
3	353	18.8	597	1 C4BP_HUMAN	P04003 homo sapien
4	350.5	18.7	2039	1 CRL_HUMAN	P17927 homo sapien
5	320.5	17.1	610	1 C4BP_BOVIN	O28065 bos taurus
6	317	16.9	558	1 C4BP_MOUSE	O63514 rattus norv
7	311.5	16.6	469	1 C4BP_MOUSE	P08607 mus musculu
8	307	16.4	1033	1 CR2_HUMAN	P20023 homo sapien
9	303.5	16.2	1025	1 CR2_MOUSE	P19070 mus musculu
10	303	16.2	661	1 F13B_HUMAN	P05160 homo sapien
11	301.5	16.1	263	1 VCP_VACCV	P10998 vaccinia vl
12	294	15.7	668	1 F13B_MOUSE	O07968 mus musculu
13	286	15.2	830	1 LEM3_HUMAN	P16109 homo sapien
14	279.5	14.9	345	1 APOH_MOUSE	O01339 mus musculu
15	279	14.9	610	1 LEM2_HUMAN	P16581 homo sapien
16	276.5	14.7	612	1 LEM2_MOUSE	O00690 mus musculu
17	274.5	14.6	768	1 LEM3_MOUSE	O01102 mus musculu
18	274	14.6	377	1 MCP_HUMAN	P15529 homo sapien
19	274	14.6	768	1 LEM3_RAT	P98106 rattus norv
20	270.5	14.4	611	1 LEM2_CANFA	P33730 canis famli
21	269.5	14.4	360	1 CCPH_HSVSA	O01016 herpesvirlu
22	267.5	14.3	345	1 APOH_HUMAN	P02749 homo sapien
23	266	14.2	484	1 LEM2_PIG	P98110 sus scrofa
24	262.5	14.0	958	1 HIG_DROME	O09101 drosophila
25	262	13.9	381	1 DAF_HUMAN	P08174 homo sapien
26	261.5	13.9	345	1 APOH_BOVIN	P17690 bos taurus
27	259	13.8	769	1 LEM3_SHEEP	P98109 ovis aries
28	252	13.4	507	1 DAF_PONPY	P49457 pongo pygma
29	246.5	13.1	507	1 DAF_CANVO	O60401 cavia porce
30	242.5	12.9	345	1 APOH_CANFA	P33703 canis famli
31	241	12.8	549	1 LEM2_RAT	P98105 rattus norv
32	239.5	12.8	551	1 LEM2_RABIT	P27113 oryctolagus
33	235	12.5	390	1 DAF1_MOUSE	O61475 mus musculu

## ALIGNMENTS

RESULT 1	ID	CFAH_HUMAN	STANDARD:	PRT: 1231 AA.	
AC	P08603	01-AUG-1988 (Rel. 08, Created)			
DT	01-JAN-1990 (Rel. 13, Last sequence update)				
DT	30-MAY-2000 (Rel. 39, Last annotation update)				
DE	COMPLEMENT FACTOR H PRECURSOR.				
GN	HFI OR HF OR CFH.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
LN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=LIVER:				
RX	MEDLINE: 86134059.				
RA	Ripoché J., Day A.J., Harris T.J.R., Sim R.B.:				
RT	"The complete amino acid sequence of human complement factor H.;"				
RL	Biochem. J. 249:593-602(1988).				
LN	[2]				
RP	SEQUENCE OF 53-445 FROM N.A.				
RX	MEDLINE: 87054207.				
RA	Schultz T.F., Schwaebler W., Stanley K.K., Weiss E., Dierich M.P.;				
RT	"Human complement factor H: Isolation of cDNA clones and partial cDNA				
RT	sequence of the 38-kDa tryptic fragment containing the binding site				
RT	for C3b.;"				
RL	Eur. J. Immunol. 16:1351-1355(1986).				
LN	[3]				
RP	SEQUENCE OF 226-445 FROM N.A., AND PARTIAL SEQUENCE.				
RX	MEDLINE: 86169701.				
RA	Kristensen T., Wetzel R.A., Tack B.F.;				
RT	"Structural analysis of human complement protein H: homology with C4b				
RT	binding protein, beta 2-glycoprotein I, and the Ba fragment of B2.;"				
RL	J. Immunol. 136:3407-3411(1986).				
LN	[4]				
RP	SEQUENCE OF 1047-1231 FROM N.A.				
RX	MEDLINE: 91201892.				
RA	Esteller C., Koistinen V., Schwaebler W., Dierich M.P., Weiss E.H.;				
RT	"Cloning of the 1.4-kb mRNA species of human complement factor H				
RT	reveals a novel member of the short consensus repeat family related				
RT	to the carboxy terminal of the classical 150-kDa molecule.;"				
RL	J. Immunol. 146:3190-3196(1991).				
LN	[5]				
RP	SEQUENCE OF 19-35.				
RX	MEDLINE: 83048213.				
RA	Sim R.B., Discolio R.G.;				
RT	"Purification and structural studies on the complement-system control				
RT	protein beta 1H (factor H).;"				
RL	Biochem. J. 205:285-293(1982).				
LN	[6]				
RP	STRUCTURE BY NMR OF 927-985 (SUSHI 16).				
RX	MEDLINE: 91278097.				
RA	Norman D.G., Barlow P.N., Baron M., Day A.J., Sim B., Campbell I.D.;				
RT	"Three-dimensional structure of a complement control protein module				
RT	in solution.;"				
RL	J. Mol. Biol. 219:717-725(1991).				

34	233.5	12.4	485	1 LEM2_BOVIN	P98107 bos taurus
35	222.5	11.9	330	1 CPH1_HUMAN	O03591 homo sapien
36	222.5	11.9	331	1 CPH2_HUMAN	O02985 homo sapien
37	222.5	11.9	407	1 DAF2_MOUSE	O61476 mus musculu
38	222	11.6	646	1 LEM3_BOVIN	P42201 bos taurus
39	217	11.6	297	1 APOH_RAT	P26644 rattus norv
40	210	11.2	372	1 LEM1_MOUSE	P18337 mus musculu
41	203.5	10.8	372	1 LEM1_HUMAN	P14151 homo sapien
42	203.5	10.8	372	1 LEM1_PANTR	O95237 pan troglod
43	202.5	10.8	372	1 LEM1_PONPY	O95235 pongo pygma
44	200	10.7	372	1 LEM1_RAT	P30836 rattus norv
45	194	10.3	372	1 LEM1_MACMU	O95198 macaca mula

FT	DISULEID	146	192		BY SIMILARITY.
FT	DISULEID	178	205		BY SIMILARITY.
FT	DISULEID	210	251		BY SIMILARITY.
FT	DISULEID	237	262		BY SIMILARITY.
FT	DISULEID	267	309		BY SIMILARITY.
FT	DISULEID	294	320		BY SIMILARITY.
FT	DISULEID	325	374		BY SIMILARITY.
FT	DISULEID	357	385		BY SIMILARITY.
FT	DISULEID	389	431		BY SIMILARITY.
FT	DISULEID	416	442		BY SIMILARITY.
FT	DISULEID	448	494		BY SIMILARITY.
FT	DISULEID	477	505		BY SIMILARITY.
FT	DISULEID	509	553		BY SIMILARITY.
FT	DISULEID	536	564		BY SIMILARITY.
FT	DISULEID	569	611		BY SIMILARITY.
FT	DISULEID	597	623		BY SIMILARITY.
FT	DISULEID	630	673		BY SIMILARITY.
FT	DISULEID	659	684		BY SIMILARITY.
FT	DISULEID	691	733		BY SIMILARITY.
FT	DISULEID	719	744		BY SIMILARITY.
FT	DISULEID	753	792		BY SIMILARITY.
FT	DISULEID	781	803		BY SIMILARITY.
FT	DISULEID	811	853		BY SIMILARITY.
FT	DISULEID	839	864		BY SIMILARITY.
FT	DISULEID	870	915		BY SIMILARITY.
FT	DISULEID	901	926		BY SIMILARITY.
FT	DISULEID	931	973		BY SIMILARITY.
FT	DISULEID	959	984		BY SIMILARITY.
FT	DISULEID	989	1032		BY SIMILARITY.
FT	DISULEID	1018	1043		BY SIMILARITY.
FT	DISULEID	1048	1091		BY SIMILARITY.
FT	DISULEID	1077	1102		BY SIMILARITY.
FT	DISULEID	1109	1152		BY SIMILARITY.
FT	DISULEID	1138	1163		BY SIMILARITY.
FT	DISULEID	1167	1218		BY SIMILARITY.
FT	DISULEID	1201	1228		BY SIMILARITY.
FT	CARBOHYD	529	529		N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	718	718		N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	802	802		N-LINKED (GLCNAC. .) .
FT	CARBOHYD	822	822		N-LINKED (GLCNAC. .) .
FT	CARBOHYD	882	882		N-LINKED (GLCNAC. .) .
FT	CARBOHYD	911	911		N-LINKED (GLCNAC. .) .
FT	CARBOHYD	1029	1029		N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1095	1095		N-LINKED (GLCNAC. .) (POTENTIAL).
FT	VARIANT	402	402		H -> Y.
FT	CONFLICT	21	21		/FTId=VAR_001979.
FT	CONFLICT	30	30		C -> Q (IN REF. 3).
FT	CONFLICT	34	34		T -> V (IN REF. 3).
FT	CONFLICT	53	54		T -> Q (IN REF. 3).
FT	CONFLICT	870	871		RP -> TL (IN REF. 2).
FT	STRAND	876	876		
FT	TURN	877	878		
FT	STRAND	879	881		
FT	STRAND	890	891		
FT	STRAND	893	894		
FT	TURN	896	900		
FT	STRAND	903	904		
FT	STRAND	907	907		
FT	STRAND	912	916		
FT	TURN	917	918		
FT	STRAND	919	920		
FT	STRAND	926	926		
SO	SEQUENCE	1231 AA;	139125 MW;	C6SEC8CF800B3FD CRC64;	
<hr/>					
Query Match                      100.0%; Score 1876; DB 1; Length 1231;					
Best local Similarity          100.0%; Pred. No. 6.9e-137;					
Matches 329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
OY	1 EDCNCLPPRRNTIELTGWSDDTYEEGTQAIYKCRPGVRSGLGNVLWVCRKGEWALPLR 60				
DB	19 ECNCNLPPRRNTIELTGSWSDDTYEEGTQAIYKCRPGVRSGLGNVLWVCRKGEWALPLR 78				



```

FT CARBOHYD 676 676 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 721 721 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 773 773 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 801 801 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1030 1030 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1061 1061 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1225 1225 N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE 1234 AA; 139082 MW; C5AC02F341B957F7 CRC64;

Query Match 67.2%; Score 1261; DB 1; Length 1234;
Best Local Similarity 65.7%; Pred. No. 1.4e-89;
Matches 216; Conservative 37; Mismatches 76; Indels 0; Gaps 0;

QY 1 EDCNELPRRNRNLTGMSDQTYPPGTOAIYCRPGYSLGNVIMVCRKGEVVALNPLR 60
DB 19 EDCKGPPEPNSLISGWSSEQLYPCPTQATYKCRPGYRLTGIYVCKNGKMWASNSR 78
QY 61 KCKRRCGHPDPTFGTFTLTGNGVEGYKAVYTCNNEGOLLGEINRCDTDGWTNDI 120
DB 79 ICRKRCGHPDPTFGSFRFRLAVGSOFEFGAKVYTCDDGQLGELIDYRCGADGWTNDI 138
QY 121 PICEVVKCLPYTAPENGKIVSSAMEPDRYHFGQAVFYCNSSGYKIEGDEMHCSDDGFV 180
DB 139 PICEVVKCLPYTELENGRIYSGAETDQGYEYGVVREFCNSGFKIEGKHEHCSEGLW 198
QY 181 SKEKRCVETISCKSPYVINGSPISOKIITKENRFQYKCNMGYEYSEGDVAVTESGWRP 240
DB 199 SNEKPCVETILTTPRVENGDDGINVRYKENERHYKCKHGVPERKGAAGVATGSGWS 258
QY 241 LSCERKSCDNYPIPGDYSPLRIKRTGDETTYOCCRNFGYPATRNTAKTCTGWTNPAP 300
DB 259 QPCEERKSPYIINGITPHRIIHRSDERIEYCNFYPTVGTSTVSKCTPTGWTNPV 318
QY 301 RCTLRPCDPTDIKHGGLYHNNRRPYPV 329
DB 319 RCTLRPCDPTDIKHGGLYHNNRRPYPV 347

RESULT 3
C4BP_HUMAN STANDARD; PRT: 597 AA.
AC P04003;
DT 23-OCT-1986 (Rel. 02, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE C4B-BINDING PROTEIN ALPHA CHAIN PRECURSOR (PROLINE-RICH PROTEIN)
DE (PRT).
GN C4BP OR C4BP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER.
RX MEDLINE: 90073699.
RA Matsuguchi T., Okamura S., Aso T., Sata T., Niho Y.;
RT "Molecular cloning of the cDNA coding for proline-rich protein (PRP):
RT identity of PRP as C4b-binding protein.";
RL Biochem. Biophys. Res. Commun. 165:138-144(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 9113199.
RA Aso T., Okamura S., Matsuguchi T., Sakamoto N., Sata T., Niho Y.;
RT "Genomic organization of the alpha chain of the human C4b-binding
RT protein gene.";
RL Biochem. Biophys. Res. Commun. 174:222-227(1991).
RN [3]
RP SEQUENCE OF 9-81 FROM N.A.
RX MEDLINE: 88242821.
RA Lintin S.J., Reid K.B.M.;
RT "Derivation of the sequence of the signal peptide in human
RT C4b-binding protein and interspecies cross-hybridisation of the C4bp

```

```

RT cDNA sequence.";
RL FEBS Lett. 232:328-332(1988).
RN [4]
RP SEQUENCE OF 203-288 FROM N.A.
RX MEDLINE: 86301119.
RA Lintin S.J., Reid K.B.M.;
RT "Studies on the structure of the human C4b-binding protein gene.";
RL FEBS Lett. 204:77-81(1986).
RN [5]
RP SEQUENCE OF 80-597 FROM N.A.
RX MEDLINE: 86025405.
RA Chung L.P., Bentley D.R., Reid K.B.M.;
RT "Molecular cloning and characterization of the cDNA coding for C4b-
RT binding protein, a regulatory protein of the classical pathway of the
RT human complement system.";
RN Biochem. J. 230:133-141(1985).
RN [6]
RP SEQUENCE OF 49-88.
RX MEDLINE: 85296001.
RA Chung L.P., Gagnon J., Reid K.B.M.;
RT "Amino acid sequence and studies of human C4b-binding protein: N-terminal
RT sequence analysis and alignment of the fragments produced by limited
RT proteolysis with chymotrypsin and the peptides produced by cyanogen
RT bromide treatment.";
RN Mol. Immunol. 22:427-435(1985).
RN [7]
RP ELECTRON MICROSCOPY, 3-DIMENSIONAL STRUCTURE, AND LIGAND BINDING.
RX MEDLINE: 83221615.
RA Dahlback B., Smith C.A., Mellier-Bernard H.J.;
RT "Visualization of human C4b-binding protein and its complexes with
RT vitamin K-dependent protein S and complement protein C4b.";
RL Proc. Natl. Acad. Sci. U.S.A. 80:3461-3465(1983).
RN [8]
RP FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
RN ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR
RN (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT
RN ALSO ACCELERATES THE DEGRADATION OF THE C4B2A COMPLEX (C3
RN CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA
RN CHAIN BINDS C4B. IT INTERACTS ALSO WITH ANTICOGULANT PROTEIN S
RN AND WITH SERUM AMYLOID P COMPONENT.
RN [9]
RP SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS
RN OF 3 POSSIBLE SORTS: A 570 KDA COMPLEX OF 7 ALPHA CHAINS AND 1
RN BETA CHAIN, A 530 KDA HOMOPENTAMER OF ALPHA CHAINS OR A 500 KDA
RN COMPLEX OF 6 ALPHA CHAINS AND 1 BETA CHAIN. THE CENTRAL BODY OF
RN THE ALPHA CHAIN HOMOPOLYMER SUPPORTS TENTACLES, EACH WITH THE
RN BINDING SITE FOR C4B AT THE END.
RN [10]
RP TISSUE SPECIFICITY: CHYLOMICRONS IN THE PLASMA.
RN [11]
RP SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.
RN [12]
RP SIMILARITY: IT IS UNCERTAIN WHETHER MET-1 OR MET-17 IS THE INITIATOR.
RN [13]
RP THIS SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial
RN entities requires a license agreement (see http://www.isb-sib.ch/announce/
RN or send an email to license@sib-sib.ch).
RN [14]
RP EMBL: M31452; AAA36507.1; -
RN EMBL: M62486; AAA36506.1; -
RN EMBL: M62475; AAA36506.1; JOINED.
RN EMBL: M62476; AAA36506.1; JOINED.
RN EMBL: M62477; AAA36506.1; JOINED.
RN EMBL: M62478; AAA36506.1; JOINED.
RN EMBL: M62479; AAA36506.1; JOINED.
RN EMBL: M62480; AAA36506.1; JOINED.
RN EMBL: M62481; AAA36506.1; JOINED.
RN EMBL: M62482; AAA36506.1; JOINED.
RN EMBL: M62484; AAA36506.1; JOINED.
RN EMBL: M62485; AAA36506.1; JOINED.
RN EMBL: X07853; CAA30701.1; -
RN EMBL: X04284; CAA51244.1; -
RN EMBL: X04296; CAA27839.1; -

```

DR EMBL: X02865; CAA26617.1; -  
 DR PIR: A33568; NBHUC4.  
 DR HSSP: P10998; 1VVC.  
 DR MIM: 120830; -  
 DR INTERPRO: IPR000436; -  
 DR PIRAM: PF00084; sushi; 8.  
 KW Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal;  
 KW Polymorphism.  
 FT SIGNAL 1 48  
 FT CHAIN 49 597  
 FT DOMAIN 49 539  
 FT REPEAT 49 109  
 FT REPEAT 112 171  
 FT REPEAT 174 235  
 FT REPEAT 238 295  
 FT REPEAT 298 361  
 FT REPEAT 364 423  
 FT REPEAT 425 481  
 FT REPEAT 483 539  
 FT DISULFID 50 96  
 FT DISULFID 81 108  
 FT DISULFID 113 154  
 FT DISULFID 140 170  
 FT DISULFID 175 217  
 FT DISULFID 203 234  
 FT DISULFID 239 281  
 FT DISULFID 267 294  
 FT DISULFID 299 348  
 FT DISULFID 332 360  
 FT DISULFID 364 387  
 FT DISULFID 7365 7409  
 FT DISULFID 7399 7422  
 FT DISULFID 426 468  
 FT DISULFID 454 480  
 FT DISULFID 484 525  
 FT DISULFID 511 538  
 FT DISULFID 546 546  
 FT DISULFID 558 558  
 FT CARBOHYD 221 221  
 FT CARBOHYD 506 506  
 FT CARBOHYD 528 528  
 FT VARIANT 92 92  
 FT VARIANT 357 357  
 FT VARIANT 357 357  
 FT SEQUENCE 597 AA; 67033 MW; 67E03F2EA85A16DD CRC64;  
 Query Match 18.8%; Score 353; DB 1; Length 597;  
 Best Local Similarity 29.38; Pred. No. 4.6e-20;  
 Matches 100; Conservative 50; Mismatches 141; Indels 50; Gaps 21;  
 20 SDQTYPESTQAIYKCRPGY-RSLGNVIMVCRK-GEWALNPLKCKRPGHGDTPPGT 77  
 67 TETRKTGTITLKYCLPEYVSHSTQGLTCSNDEWV-YNFE-CIYKCRHPRGE----- 118  
 78 FTLLGNV-----FEYKAVYITNEGQGLGEINRECDID---GWTNDIPICEVYKC 128  
 119 -LRNGQVEIKTIDLSFGQISEGSEGFLLIGSTTSR-CEVQDRKGVGSHPLPQCEIVKC 175  
 129 LPVAPRENGKIYSSAMEPDREYHFGQAVFVNCNGYKIEGDEMHCSD---GFWSKRK 184  
 176 KPPPIRNGRHSGE-----ENFYAGFSVYISCDPRFSLGHAISICVYVNETIGVWRPSP 231  
 185 PKVVEISCKSPDVINGSPISQ-KIYKENERFQYKCNMGYSESGDAV--C-TESGWRP 240  
 232 PTCETIKCRKPKPVSHGEWVGSGFGPIYKNTIVFCQKGFVL--RGSSVIHCDADSKMNP 289  
 241 L-PSEEEKSCDN-PYIPNDYS--PLRIK---HRTGDEITTCGRNGFYAT--RGNTAKC- 291  
 290 SPPACEPNSCINLPDIPHASWETYPRTKEDVYVGVYALRCHPGYKPTTDEPTTIVICQ 349  
 292 TSTGWIPARCTLKPQDYPDIKHGLYHENMRRP-----YF 327

DR 350 KNLRWTPYGCALCCPEPKLNNGEITQHKRSRANHCYVF 390  
 RESULT 4  
 CRL\_HUMAN  
 ID CRL\_HUMAN STANDARD; PRT; 2039 AA.  
 AC P17927;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE COMPLEMENT RECEPTOR TYPE 1 PRECURSOR (C3B/C4B RECEPTOR) (CD35 ANTIGEN).  
 GN CRL OR C3BR.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 89035992.  
 RA Klickstein L.B., Bartow T.J., Miletic V., Rabson L.D., Smith J.A., Fearon D.T.;  
 RT "Identification of distinct C3b and C4b recognition sites in the human C3b/C4b receptor (CRL, CD35) by deletion mutagenesis.";  
 RL J. Exp. Med. 168:1699-1717(1988).  
 RN [2]  
 RP SEQUENCE OF 503-2039 FROM N.A.  
 RX MEDLINE: 87168191.  
 RA Klickstein L.B., Wong W.W., Smith J.A., Weis J.H., Wilson J.G., Fearon D.T.;  
 RT "Human C3b/C4b receptor (CRL). Demonstration of long homologous repeating domains that are composed of the short consensus repeats characteristics of C3/C4 binding proteins.";  
 RL J. Exp. Med. 165:1095-1112(1987).  
 RN [3]  
 RP SEQUENCE OF 761-783; 831-845 AND 1179-1195 FROM N.A.  
 RX MEDLINE: 86067975.  
 RA Wong W.W., Klickstein L.B., Smith J.A., Weis J.H., Fearon D.T.;  
 RT "Identification of a partial cDNA clone for the human receptor for complement fragments C3b/C4b.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:7711-7715(1985).  
 CC -!- FUNCTION: CRL, PRESENT ON ERYTHROCYTES, LEUCOCYTES, GLOMERULAR PODOCYTES, AND SPLENIC FOLLICULAR DENDRITIC CELLS. MEDIATES THE BINDING BY THESE CELLS OF PARTICLES AND IMMUNE COMPLEXES THAT HAVE ACTIVATED COMPLEMENT.  
 CC -!- SUBUNIT: MONOMER.  
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -!- POLYMORPHISM: CRL IS RESPONSIBLE FOR THE KNOX BLOOD GROUP SYSTEM.  
 CC -!- MISCELLANEOUS: SEVEN SHORT CONSENSUS REPEATS (SCR) CONSTITUTE A LONG HOMOLOGOUS REPEAT (LHR). THE N-TERMINAL TWO SRCS OF LHR-A CONTAINED A SITE DETERMINING C4 SPECIFICITY, AND THE N-TERMINAL TWO SRCS OF LHR-B AND -C EACH HAD A SITE DETERMINING C3 SPECIFICITY.  
 CC -!- SIMILARITY: THIS IS THE SEQUENCE OF THE F ALLOTYPE OF CRL.  
 CC -!- SIMILARITY: CONTAINS 30 SUSHI (SCR) REPEATS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 DR EMBL: M11569; AAA52297.1; -  
 DR EMBL: M11617; AAA52298.1; -  
 DR EMBL: M11618; AAA52299.1; -  
 DR EMBL: Y00816; CAA68755.1; -  
 DR EMBL: X05309; CAA28933.1; -  
 DR PIR: A28507; A28507.  
 DR PIR: A24748; A24748.  
 DR PIR: B24748; B24748.  
 DR PIR: C24748; C24748.



FT	REPEAT	486	542		SUSHI 8.	BY SIMILARITY.
FT	DISULFID	50	95		BY SIMILARITY.	
FT	DISULFID	80	107		BY SIMILARITY.	
FT	DISULFID	112	153		BY SIMILARITY.	
FT	DISULFID	139	169		BY SIMILARITY.	
FT	DISULFID	174	217		BY SIMILARITY.	
FT	DISULFID	203	234		BY SIMILARITY.	
FT	DISULFID	239	261		BY SIMILARITY.	
FT	DISULFID	267	294		BY SIMILARITY.	
FT	DISULFID	299	350		BY SIMILARITY.	
FT	DISULFID	334	362		BY SIMILARITY.	
FT	DISULFID	7366	7390		BY SIMILARITY.	
FT	DISULFID	7367	7412		BY SIMILARITY.	
FT	DISULFID	7402	7425		BY SIMILARITY.	
FT	DISULFID	429	471		BY SIMILARITY.	
FT	DISULFID	457	483		BY SIMILARITY.	
FT	DISULFID	487	528		BY SIMILARITY.	
FT	DISULFID	514	541		BY SIMILARITY.	
FT	DISULFID	549	549		INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).	
FT	DISULFID	561	561		INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).	
FT	CARBOHYD	66	66		N-LINKED (GLCNAC. . . ) (POTENTIAL).	
FT	CARBOHYD	221	221		N-LINKED (GLCNAC. . . ) (POTENTIAL).	
FT	CARBOHYD	525	525		N-LINKED (GLCNAC. . . ) (POTENTIAL).	
FT	CARBOHYD	602	602		N-LINKED (GLCNAC. . . ) (POTENTIAL).	
SO	SEQUENCE	610 AA;	68886 MW;	D806B270E8A0658 CRC64;		
<hr/>						
Query Match		17.1%;	Score 320.5;	DB 1;	Length 610;	
Best Local Similarity		27.7%;	Pred. No. 1.5e-17;			
Matches 99;		Conservative 54;	Mismatches 142;	Indels 63;	Gaps	
<hr/>						
QY	6 LPPRRNTEILLGMSDDQTPEGCDAIKKCPGRV--SLGVNIWICRKGAVNALPLRKQ	63				
Db	52 IPLLDPAPFINELNEFRFGTTLRTCTRGRISSRKNFLIDGDIDNM---KYKFCV	108				
QY	64 KRPGHGFDFEFEGFTLTGCVN-----FEYGKVAIVVICNEGOLLGEIN-YRECDDTG-	115				
Db	109 KKRENGE-----LLNGOVIKTDTYSFSGEIEFSCSEGYVLIGANSYCOLODKGV	161				
QY	116 WTNDIPICEVVKCLPYTAPENGKIYVSAMERDRYHNHGQAVRPVCNSGKYTEDEENHS	175				
Db	162 WSDPLPOCIIKKCEPPTISNGR--HHGGDED-FYTGGSVTYSCDHFSLGKAASISCR	218				
QY	176 DD----GWMSKEPKRCVASICSPDIVINGSPIISKI-----LYKENREFQKCMNYEY	225				
Db	219 VENKTIGWMSPPSCKKVITGVGVADG-----KITSGERPIITYQOSTIYAACKNGFRL	273				
QY	226 SERGDV--C-TESGWR-PLPSCEEKSC-DNPVTPNGDYSPLRILKHRTGDE-----	271				
Db	274 --EGDSLTHCBADNSWNPNPPTCELNGCLGIPLHPHALME--RYDHOTQEQQVYDIGFV	329				
QY	272 ITTCRNRFYATNG-NIAKTSST-GWIAPARCLLPCDDYDIKHGG--LYHEMMRRP	325				
Db	330 LSYKFHFYEKPETDPTVTTCOSLNLEMSPTYECKEVCCPEPLNNYGSITLIH---RRP	384				
<hr/>						
RESULT 6						
C4BP_RAT		STANDARD;	PRT;	558 AA.		
ID	C4BP_RAT					
AC	O63514;					
DT	01-NOV-1997 (Rel. 35, Created)					
DT	01-NOV-1997 (Rel. 35, Last sequence update)					
DT	15-JUL-1998 (Rel. 36, Last annotation update)					
DE	C4B-BINDING PROTEIN ALPHA CHAIN PRECURSOR.					
GN	C4BA.					
OS	Rattus norvegicus (Rat).					
OC	Eumetazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN-SPAGUE-DAWLEY; TISSUE=LIVER;					
RX	MEDLINE: 97166087.					
RA	Hillarp A., Wiklund H., Thern A., Dahlback B.;					

RT "Molecular cloning of rat C4b binding protein alpha- and beta-chains: structural and functional relationships among human, bovine, rabbit, mouse, and rat proteins." J. Immunol. 158:1315-1323(1997).

CC -1- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT ALSO ACCELERATES THE DEGRADATION OF THE C4B2A COMPLEX (C3 CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA CHAIN BINDS C4B. IT INTERACTS ALSO WITH ANTICOAGULANT PROTEIN S AND WITH SERUM AMYLOID P COMPONENT.

CC -1- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS.

CC -1- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.

CC -1- SIMILARITY: CONTAINS 8 SUSHI (SCR) REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: Z50051; CAA90391.1; .

DR HSSP: P10998; IYVC.

DR INTERPRO: IPR000436; .

DR PFAM: PF00084; sushi; 8.

KW Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal.

FT CHAIN 1 13 BY SIMILARITY.

FT SIGNAL 1 13 C4B-BINDING PROTEIN ALPHA CHAIN.

FT DOMAIN 14 558 8 X SUSHI (SCR) REPEATS.

FT REPEAT 14 502 SUSHI 1.

FT REPEAT 14 73 SUSHI 2.

FT REPEAT 138 135 SUSHI 3.

FT REPEAT 203 259 SUSHI 4.

FT REPEAT 262 325 SUSHI 5.

FT REPEAT 328 387 SUSHI 6.

FT REPEAT 389 444 SUSHI 7.

FT REPEAT 446 502 SUSHI 8.

FT DISULFID 15 60 BY SIMILARITY.

FT DISULFID 45 72 BY SIMILARITY.

FT DISULFID 77 118 BY SIMILARITY.

FT DISULFID 104 134 BY SIMILARITY.

FT DISULFID 139 182 BY SIMILARITY.

FT DISULFID 168 199 BY SIMILARITY.

FT DISULFID 204 246 BY SIMILARITY.

FT DISULFID 232 258 BY SIMILARITY.

FT DISULFID 263 312 BY SIMILARITY.

FT DISULFID 296 324 BY SIMILARITY.

FT DISULFID 328 351 BY SIMILARITY.

FT DISULFID 329 373 BY SIMILARITY.

FT DISULFID 363 386 BY SIMILARITY.

FT DISULFID 390 431 BY SIMILARITY.

FT DISULFID 417 443 BY SIMILARITY.

FT DISULFID 447 488 BY SIMILARITY.

FT DISULFID 474 501 BY SIMILARITY.

FT DISULFID 509 521 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).

FT DISULFID 521 521 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).

FT CARBOHYD 31 31 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 177 177 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 491 491 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 558 AA; 62266 MW; 592F0C667ED1E5FF CRC64;

Query Match 16.9%; Score 317; DB 1; Length 558;  
Best Local Similarity 27.0%; Pred. No. 2.3e-17;  
Matches 110; Conservative 46; Mismatches 128; Indels 124; Gaps 24;

QY 7 PPRNTEILGWSMDQT-YPEGTOAIYKCRPGY-RSLGNVIMVCRK-GEW---VALNLR 60  
DB 17 PPDPLPALPASEMNQIDFESHITLRINCRCRGYSRASSQSGLYCKPKQMOINIA----- 71

QY 61 KCKRPGHGDTPFGFTLTGNNV-----FEYKAVAVYCNNGYOLG-EIIVRECDTD 114  
DB 72 -CKKSGRNNGD-----LQNKVEYKIDFLGSOIEFSCSGYILIGSTSYCEIQK 123

QY 115 G--WINDIPICEVKKLPYAPENGK-----IYSSAMEPD 147  
DB 124 GVSMDPLPECVIAKCMGPDIDNGKHNGREEFYRSVTKCPDDFTLLGNASTITG 183

QY 148 -----REYFGAVPFVNSGKIGDEE 171  
DB 184 VVKITGVWSPSPPTCEIICPPKVLHGHTNSGFHFKYKDSVRFVQKGFVLRSVG 243

QY 172 MHCSDDGFWSKPKCVETSCKS-PRVINGSPIS-----OKIYKNERFQYCNNGY 225  
DB 244 IHCADGSMSPVYVCELNSCTDIPIDNALITSPRKEDYIPVGYLARIYCRGYE 302

QY 226 SERG--DAVCTES-GWRPLPSCHEKSCDNPYIPNGYSPLR-TKHRT-----GD 270  
DB 303 ATRQPMTVICQKDLWSMRLGCKEICCPVP-----DPKSVRIQHKAPHDNDCTYF 357

QY 271 EITYOCRMGEFYPATRGNTAKCTSTG-WIP-APRCITLKPDDYD-IKHG 315  
DB 358 EYSYIQNDIM-----LTATCKSDGTWHPRTPSCH-QSCDFPALAHG 399

RESULT 7  
C4BP\_MOUSE STANDARD; PRT; 469 AA.  
ID C4BP\_MOUSE P08607:  
AC 01-AUG-1988 (rel. 08, Created)  
DT 01-FEB-1994 (rel. 28, Last sequence update)  
DT 01-NOV-1997 (rel. 35, Last annotation update)  
DE C4B-BINDING PROTEIN PRECURSOR (C4BP).  
GN C4BPA OR C4BP.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 88024997.  
RA Kristensen T., Ogata R.T., Chung L.P., Reid K.B.M., Jack B.F.;  
RT "cDNA structure of murine C4b-binding protein, a regulatory component of the serum complement system.";  
RL Biochemistry 26:468-467(1987).

CC -1- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT ALSO ACCELERATES THE DEGRADATION OF THE C4B2A COMPLEX (C3 CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA CHAIN BINDS C4B. IT INTERACTS ALSO WITH SERUM AMYLOID P COMPONENT.

CC -1- SUBUNIT: HOMOPOLYMER; NOT COVALENTLY LINKED. MOUSE LACKS THE BETA CHAIN OF C4BP.

CC -1- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.

CC -1- SIMILARITY: CONTAINS 6 SUSHI (SCR) REPEATS.

CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-44 IS THE INITIATOR.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: M17122; AAA37312.1; ALT\_INIT.  
DR PIR: A27117; NEMSC4.  
DR HSSP: P10998; IYVC.  
DR MGD: MGI:88229; C4BP.  
DR INTERPRO: IPR000436; .  
DR PFAM: PF00084; sushi; 6.  
KW Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal.

```

FT CHAIN 57 469 CAB-BINDING PROTEIN.
FT DOMAIN 57 414 6 X SUSHI (SCR) REPEATS.
FT REPEAT 57 116 SUSHI 1.
FT REPEAT 119 177 SUSHI 2.
FT REPEAT 180 241 SUSHI 3.
FT REPEAT 244 300 SUSHI 4.
FT REPEAT 302 356 SUSHI 5.
FT REPEAT 358 413 SUSHI 6.
FT DISULFID 58 103 BY SIMILARITY.
FT DISULFID 88 115 BY SIMILARITY.
FT DISULFID 120 160 BY SIMILARITY.
FT DISULFID 146 176 BY SIMILARITY.
FT DISULFID 181 223 BY SIMILARITY.
FT DISULFID 209 240 BY SIMILARITY.
FT DISULFID 245 287 BY SIMILARITY.
FT DISULFID 273 299 BY SIMILARITY.
FT DISULFID 303 343 BY SIMILARITY.
FT DISULFID 329 355 BY SIMILARITY.
FT DISULFID 359 400 BY SIMILARITY.
FT DISULFID 386 413 BY SIMILARITY.
FT CARBOHYD 74 74 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 275 275 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 292 292 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 366 366 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 381 381 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 428 428 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 469 AA; 51551 MW; 41E137CB8D8C6321 CRC64;

```

```

Query Match 16.6%; Score 311.5; DB 1; Length 469;
Best Local Similarity 28.5%; Pred. No. 5,5e-17;
Matches 94; Conservative 54; Mismatches 149; Indels 33; Gaps 19;

```

```

QY 7 PPRNPFILGMSDOR-YPEGTOAIYKCRPGY-RSLGNVIMVCR-KGEM-VALNPLRKC 62
DB 60 PPAIPALPASADVNRKDFESHITLKECLPGYRGISRMVYKPSGEWETIS----C 115
QY 63 QKRPCHGHPGDTPEGTFTLTGNGVPEYGVKAVYTCNEGYYOLLGEINRECDTDG---WTN 118
DB 116 AKKRCHRPGLDNG---YVNGETITFGSQIEFSCQEGFLLVGS-STSCSEVRGKVAWSN 171
QY 119 DIPCEVYKCLPTAPENKGVSSAMPEDRYHGOAVRFNCNSGKYEGBEMHCS--- 175
DB 172 PFPECVIVKCGPPDINGK--HSGTDFEYFYNHG--ISYTCDFGLVGSPIGCTVYN 227
QY 176 -DDGFWKKEKPKVEISCKSPDIVNGSPIS-OKIITYKENERFOYKCNMGYSEKGDVAVC 233
DB 228 KTVYVWSSSPPTCKEIKISQPNIIHGIVYSGKATYTHRDSVRLACGLNGYVLRGHNVEC 287
QY 234 TESSG-WRPLPSCEKSCD-NFYIPNGDYSPLRIKHRTGDEITTYOCRNFGYPATKGNIAKC 291
DB 288 QGNGMWSLPTGC-DEDCDLPAIVNGYVTSWVYSKIT--LVTYEDCKG-YRLVKAIIISC 343
QY 292 TSTGWT-PAPRCTLKPCDYPIKHHGLYHE 320
DB 344 SFSKWKGTAPCKAL-CQKPEVNGTISDE 372

```

```

RESULT 8
CR2_HUMAN STANDARD; PRT; 1033 AA.
AC P20023;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE COMPLEMENT RECEPTOR TYPE 2 PRECURSOR (CR2) (COMPLEMENT C3D RECEPTOR)
DE (EPSTEIN-BARR VIRUS RECEPTOR) (CD21 ANTIGEN).
GN CR2 OR C3DR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]

```

```

RP SOURCE FROM N.A.
RX MEDLINE; 89123277.
RA Fujisaku A., Harley J.B., Frank M.B., Gruner B.A., Frazier B.,
RA Holers V.M.;
RT "Genomic organization and polymorphisms of the human C3d/Epstein-Barr
RT virus receptor.";
RL J. Biol. Chem. 264:2118-2125(1989).
[2]
RP SEQUENCE OF 226-233; 256-267; 332-341; 667-677 AND 898-908.
RX MEDLINE; 86287311.
RA Weis J.J., Fearon D.T., Klickstein L.B., Wong W.W., Richards S.A.,
RA de Bruyn Kops A., Smith J.A., Weis J.H.;
RT "Identification of a partial cDNA clone for the C3d/Epstein-Barr
RT virus receptor of human B lymphocytes: homology with the receptor for
RT fragments C3b and C4b of the third and fourth components of
RT complement.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:5639-5643(1986).
CC -1- FUNCTION: RECEPTOR FOR COMPLEMENT C3D AND FOR THE EPSTEIN-BARR
CC VIRUS ON HUMAN B-CELLS. PARTICIPATES IN B LYMPHOCYTES ACTIVATION.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: MATURE B LYMPHOCYTES AND FOLLICULAR
CC DENDRITIC CELLS OF THE SPLEEN.
CC -1- SIMILARITY: TO MOUSE CR2. CD21 IS A MEMBER OF THE REGULATORS OF
CC COMPLEMENT ACTIVATION (RCA) FAMILY.
CC -1- SIMILARITY: CONTAINS 15 SUSHI (SCR) REPEATS.
CC -1- DATABASE: NAME-PROV; NOTE-CD guide CD21 entry;
CC WWW="HTTP://WWW.NCBI.NLM.NIH.GOV/PROV/CD/CD21.HTM".
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M26004; AAA35786.1; -
DR EMBL; M26016; AAB04638.1; -
DR EMBL; M24007; AAB04638.1; JOINED.
DR EMBL; M24008; AAB04638.1; JOINED.
DR EMBL; M24009; AAB04638.1; JOINED.
DR EMBL; M24010; AAB04638.1; JOINED.
DR EMBL; M24011; AAB04638.1; JOINED.
DR EMBL; M26009; AAB04638.1; JOINED.
DR EMBL; M26010; AAB04638.1; JOINED.
DR EMBL; M26011; AAB04638.1; JOINED.
DR EMBL; M26012; AAB04638.1; JOINED.
DR EMBL; M26013; AAB04638.1; JOINED.
DR EMBL; M26014; AAB04638.1; JOINED.
DR EMBL; M26015; AAB04638.1; JOINED.
DR PIR; A32036; A32036.
DR PIR; A24319; A24319.
DR PIR; B24319; B24319.
DR PIR; C24319; C24319.
DR PIR; D24319; D24319.
DR PIR; E24319; E24319.
DR PIR; F24319; F24319.
DR HSSP; P10998; IYVC.
DR MIM; 120650; -.
DR INTERPRO: IPR000436; -.
DR PFM: PF00084; sush1.15.
KW Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
KW Receptor; Sush1.
FT SIGNAL 1 20
FT CHAIN 21 1033 COMPLEMENT RECEPTOR TYPE 2.
FT DOMAIN 21 971 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 972 999 POTENTIAL.
FT DOMAIN 1000 1033 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 22 971 15 X SUSHI (SCR) REPEATS.
FT REPEAT 22 83 SUSHI 1.
FT REPEAT 90 147 SUSHI 2.
FT REPEAT 153 211 SUSHI 3.
FT REPEAT 214 272 SUSHI 4.

```

```

FT REPEAT 275 343 SUSHI 5.
FT REPEAT 350 407 SUSHI 6.
FT REPEAT 409 467 SUSHI 7.
FT REPEAT 470 523 SUSHI 8.
FT REPEAT 526 594 SUSHI 9.
FT REPEAT 601 658 SUSHI 10.
FT REPEAT 661 715 SUSHI 11.
FT REPEAT 718 780 SUSHI 12.
FT REPEAT 787 844 SUSHI 13.
FT REPEAT 850 908 SUSHI 14.
FT REPEAT 911 969 SUSHI 15.
FT DISULFID 23 65 BY SIMILARITY.
FT DISULFID 51 82 BY SIMILARITY.
FT DISULFID 91 132 BY SIMILARITY.
FT DISULFID 118 146 BY SIMILARITY.
FT DISULFID 154 197 BY SIMILARITY.
FT DISULFID 183 210 BY SIMILARITY.
FT DISULFID 215 256 BY SIMILARITY.
FT DISULFID 242 271 BY SIMILARITY.
FT DISULFID 276 325 BY SIMILARITY.
FT DISULFID 305 342 BY SIMILARITY.
FT DISULFID 351 393 BY SIMILARITY.
FT DISULFID 379 406 BY SIMILARITY.
FT DISULFID 410 453 BY SIMILARITY.
FT DISULFID 439 466 BY SIMILARITY.
FT DISULFID 471 509 BY SIMILARITY.
FT DISULFID 495 522 BY SIMILARITY.
FT DISULFID 527 576 BY SIMILARITY.
FT DISULFID 556 593 BY SIMILARITY.
FT DISULFID 602 644 BY SIMILARITY.
FT DISULFID 630 657 BY SIMILARITY.
FT DISULFID 662 699 BY SIMILARITY.
FT DISULFID 719 762 BY SIMILARITY.
FT DISULFID 748 779 BY SIMILARITY.
FT DISULFID 788 830 BY SIMILARITY.
FT DISULFID 816 843 BY SIMILARITY.
FT DISULFID 851 894 BY SIMILARITY.
FT DISULFID 880 907 BY SIMILARITY.
FT DISULFID 912 955 BY SIMILARITY.
FT DISULFID 941 968 BY SIMILARITY.
FT CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 194 294 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 372 372 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 492 492 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 682 682 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 800 800 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 823 823 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 861 861 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 911 911 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 667 667 O -> D (IN REF. 2).
FT CONFLICT 902 902 O -> G (IN REF. 2).
FT CONFLICT 906 906 H -> L (IN REF. 2).
SQ SEQUENCE 1033 AA; 112973 MW; 1749DBA07847ADA CRC64;

```

Query Match 16.4%; Score 307; DB 1; Length 1033;  
 Best Local Similarity 23.7%; Pred. No. 3e-16;  
 Matches 97; Conservative 58; Mismatches 139; Indels 116; Gaps 22;

```

QY 1 EDCEMLPRRNTLITGMSDO---TYPEGQAIKCRPGYSLGNVIVC-RKGEWAL 56
DQ 408 KEC-QAPD-----NINLGQKEDRHVPRDPTGISKSNPGYLVGEESIQCTSEGVWP- 461
QY 57 NPLKRCQKRCPCGHRG---DIPFGTFTLTGNGVFEYGVKANYTCNEGYQLLGEINVRCD 112
DQ 462 -PVOQCKVAACEATGRQLLTRPQHGFVRPDV-----SSCGEGYKLSGSV-YQECQ 510
QY 113 -TDGWNIDPICEVVKCLPVTAPENGKIVSSAMEPDRYHFGQAVRVCNSG-----YK 165
DQ 511 CTIMFMEIRICKETTCPPVYVINGAHTGSSLF---DPIGTITVYTCNCPGPERGVEFS 567

```

```

QY 166 IEGDEMHCCSD-----GFWSEKPKC-----VEISCKSPDVINGSPISOK-IYKENERFO 216
DQ 568 LIGSTIRCTNSDOERBTWGSAPLCKLSLAVCCSHVHLANGIKISGEAPFYNDPTVT 627
QY 217 YKCMNGYEYSERGDVAVC-TESGWRP-LPSCSEKSDN----- 251
DQ 628 FKCYSGFTLKSSQIRCKADNTMDPEIPVCEKEFCQHVRSLOELPAGSRVELVNTSCOD 687
QY 252 -----PIIPMGDISPLRIKH-RIGDEITTYQ 275
DQ 688 GYOLTGAIYOMCODAENGIFPKIPLCKVIHCHPPIVYNGKHTGMAENFLYGNEVSE 747
QY 276 CRNGEYPATRGNTAKCTST-----GWI-PAPRCTLKP-----CDYPOIKHG 315
DQ 748 CDQGEY-LIGKKLQCRSDSKSGHSGSPDCLRSPVYRCNPEYKVG 796

RESULT 9
CR2_MOUSE STANDARD; PRT; 1025 AA.
AC P19070;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE COMPLEMENT RECEPTOR TYPE 2 PRECURSOR (CR2) (COMPLEMENT C3D RECEPTOR).
GN CR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RA MEDLINE; 90229735.
RA Fingerboth J.D.;
RT "Comparative structure and evolution of murine CR2. The homolog of
  the human C3d/EBV receptor (CD21).";
RL J. Immunol. 144:3458-3467(1990).
[2]
RP SEQUENCE OF 12-1025 FROM N.A.
RA MEDLINE; 91010789.
RA Molina H., Kinoshita T., Inoue K., Cappel J.C., Holers V.M.;
RT "A molecular and immunochemical characterization of mouse CR2.
  Evidence for a single gene model of mouse complement receptors 1 and
  2.";
RL J. Immunol. 145:2974-2983(1990).
[3]
RP SEQUENCE OF 343-401 AND 991-1025 FROM N.A.
RA MEDLINE; 89098890.
RA Fingerboth J.D., Benedict M.A., Levy D.N., Strominger J.L.;
RT "Identification of murine complement receptor type 2.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:242-246(1989).
[4]
RP SEQUENCE OF 289-1025 FROM N.A.
RA MEDLINE; 89381350.
RA Kurtz C.B., Paul M.S., Aegerter M., Weis J.J., Weis J.H.;
RT "Murine complement receptor gene family. II. Identification and
  characterization of the murine homolog (Cr2) to human CR2 and its
  molecular linkage to Cr1.";
RL J. Immunol. 143:2058-2067(1989).
[5]
RP FUNCTION: RECEPTOR FOR COMPLEMENT C3D. PARTICIPATES IN B
  LYMPHOCYTES ACTIVATION.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: B LYMPHOCYTES.
CC -1- SIMILARITY: TO HUMAN C3D/EBV RECEPTOR (CD21).
CC -1- SIMILARITY: CONTAINS 15 SUSHI (SCR) REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation-
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb.ch/announce/)

```

CC	or send an email to <a href="mailto:license@ib-sib.ch">license@ib-sib.ch</a> ).			
CC	-----			
DR	EMBL; M81083; AAA37451.1; -			
DR	EMBL; M35684; AAA37448.1; -			
DR	EMBL; M61132; AAA63295.1; -			
DR	EMBL; M35685; AAA37450.1; -			ALT_SEQ.
DR	EMBL; M29281; AAA37447.1; -			
DR	PIR; A43526; A43526.			
DR	HSSP; P10998; 1IYC.			
DR	MGD; MGT;88489; CR2.			
DR	INTERPRO; IPR000436; -			
DR	PFAM; PF00084; sushi1.14.			
KW	Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal; Receptor; Sushi.			
FT	SIGNAL	1	11	
FT	CHAIN	12	1025	COMPLEMENT RECEPTOR TYPE 2.
FT	DOMAIN	12	963	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	964	990	POTENTIAL.
FT	DOMAIN	991	1025	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	13	959	15 X SUSHI (SCR) REPEATS.
FT	REPEAT	13	74	SUSHI 1.
FT	REPEAT	81	139	SUSHI 2.
FT	REPEAT	145	203	SUSHI 3.
FT	REPEAT	206	264	SUSHI 4.
FT	REPEAT	267	335	SUSHI 5.
FT	REPEAT	342	399	SUSHI 6.
FT	REPEAT	401	458	SUSHI 7.
FT	REPEAT	462	515	SUSHI 8.
FT	REPEAT	518	566	SUSHI 9.
FT	REPEAT	593	649	SUSHI 10.
FT	REPEAT	653	705	SUSHI 11.
FT	REPEAT	708	769	SUSHI 12.
FT	REPEAT	777	834	SUSHI 13.
FT	REPEAT	840	898	SUSHI 14.
FT	REPEAT	901	959	SUSHI 15.
FT	DISULFID	14	56	BY SIMILARITY.
FT	DISULFID	42	73	BY SIMILARITY.
FT	DISULFID	82	124	BY SIMILARITY.
FT	DISULFID	110	138	BY SIMILARITY.
FT	DISULFID	146	189	BY SIMILARITY.
FT	DISULFID	175	202	BY SIMILARITY.
FT	DISULFID	207	248	BY SIMILARITY.
FT	DISULFID	234	263	BY SIMILARITY.
FT	DISULFID	268	317	BY SIMILARITY.
FT	DISULFID	297	334	BY SIMILARITY.
FT	DISULFID	343	385	BY SIMILARITY.
FT	DISULFID	371	398	BY SIMILARITY.
FT	DISULFID	402	445	BY SIMILARITY.
FT	DISULFID	431	458	BY SIMILARITY.
FT	DISULFID	463	501	BY SIMILARITY.
FT	DISULFID	487	514	BY SIMILARITY.
FT	DISULFID	519	568	BY SIMILARITY.
FT	DISULFID	548	585	BY SIMILARITY.
FT	DISULFID	594	636	BY SIMILARITY.
FT	DISULFID	622	649	BY SIMILARITY.
FT	DISULFID	654	689	BY SIMILARITY.
FT	DISULFID	675	704	BY SIMILARITY.
FT	DISULFID	709	752	BY SIMILARITY.
FT	DISULFID	738	769	BY SIMILARITY.
FT	DISULFID	778	820	BY SIMILARITY.
FT	DISULFID	806	833	BY SIMILARITY.
FT	DISULFID	841	884	BY SIMILARITY.
FT	DISULFID	870	897	BY SIMILARITY.
FT	DISULFID	902	945	BY SIMILARITY.
FT	DISULFID	931	958	BY SIMILARITY.
FT	CARBOHYD	77	77	N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT	CARBOHYD	113	113	N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT	CARBOHYD	276	276	N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT	CARBOHYD	316	316	N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT	CARBOHYD	364	364	N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT	CARBOHYD	380	380	N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT	CARBOHYD	484	484	N-LINKED (GLCNAC. . .) (POTENTIAL. .)
FT	CARBOHYD	527	527	N-LINKED (GLCNAC. . .) (POTENTIAL. .)

Query Match	Best Local Similarity	Score	DB 1:	Length	DB 2:	Length
Matches	96; Conservative	37; Mismatches	146; Indels	67; Gaps	22	
QY 2	DCNELPRRRTTEILTGSWSQDTYPECTQAIYKCRPGYSLGNWIMVC-RKGEWALNPLR 60	16.2%;	Score 303.5;	DB 1:	Length 1025;	
Db 145	ECSPILPTIHNGH-HTGQHVQ-QVACLSTVYSCGEYGLTGKTKIKLSSGDMQGVIP-- 200	27.7%;	Pred. No. 5.5e-16;			
QY 61	KCKRRPGCHGHDPRFGFTILTGNN-----FEYKNAVYTCNEGVLGELINREC----- 111					
Db 201	TCKEAGECHHKRP-----NGQVKEPLSLQVGTYYFSCNEGVLQDQGS-SQCIYVE 252					
QY 112	DTDEMTDPIICEYVKCLPYTAPENKIVSSAMEPDRREYHGOAVRFVCS-----GYK 165					
Db 253	QKAIWTKK-PVCKEILCLPPRPYRNSHNGSSE---NVPYSIVTYICDPSPEKVSFT 308					
QY 166	IEGDEMHCS---DDGFWSKEKPKCV-----EISCKSPVINGSPIS-QKIITYENERFQ 216					
Db 309	LIGERTKINTGSGOKTGWGPAPYCVLSTSAVLCLQPKIKRGQILSLKOSYSTNPVA 368					
QY 217	YKCMNGEYSRGDAVCTESG-WR-ELPSCSEKSCNP-----YIPNDYSPLR 263					
Db 369	FSCBERFLKGNRSIRCNNAHGTWEPYPPVC-EKGCQAPPKIINGKEDSYLNTD----- 422					
QY 264	IKHRTGDEITYOCRNGFYPATRGNTAKCTSTG-WIP-ADPCTLKPC 307					
Db 423	---PGTSIRYSCDPG-YLLVGEDTLHCTPEGKWPITPQCTVADC 463					
RESULT 10						
FL3B_HUMAN	STANDARD;	PRT;	661 AA.			
ID P05160;						
DC 13-AUG-1987 (Rel. 05, Created)						
DT 01-AUG-1980 (Rel. 15, Last sequence update)						
DT 15-JUL-1999 (Rel. 38, Last annotation update)						
DE COGULATIVE FACTOR XIII B CHAIN PRECURSOR (EC 2.3.2.13) (PROTEIN-GLUTAMINE GAMMA-GLUTAMYLTRANSFERASE B CHAIN)						
DE CHAIN).						
GN FL3B.						
OS Homo sapiens (Human).						
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.						
RN [1]						
RN SEQUENCE FROM N.A.						
RX MEDLINE; 91105054.						
RA Bottenus R.E., Ichinose A., Davie E.W.;						
RT "Nucleotide sequence of the gene for the b subunit of human factor XIII."						
RL Biochemistry 29:11195-11209(1990).						
RN [2]						
RN SEQUENCE OF 2-661 FROM N.A.						
RX MEDLINE; 87026535.						
RA Ichinose A., McMullen B.A., Fujikawa K., Davie E.W.;						
RT "Amino acid sequence of the b subunit of human factor XIII, a protein composed of ten repetitive segments."						
RL Biochemistry 25:4633-4638(1986).						

```

FT DISULFID 486 505 BY SIMILARITY.
FT DISULFID 486 515 BY SIMILARITY.
FT DISULFID 524 567 BY SIMILARITY.
FT DISULFID 524 578 BY SIMILARITY.
FT DISULFID 582 636 BY SIMILARITY.
FT DISULFID 616 646 BY SIMILARITY.
FT CARBOHYD 162 162 N-LINKED (GLCNAc . . .) (POTENTIAL).
FT CARBOHYD 545 545 N-LINKED (GLCNAc . . .) (POTENTIAL).
FT SITE 617 619 CELL ATTACHMENT SITE.
FT VARIANT 450 450 C>=F (IN F13 DEFICIENCY).
SQ SEQUENCE 661 AA; 75491 MW; 57A2FB46560857F2 CRC64;
/FTID=VAR_007475.

Query Match 16.2%; Score 303; DB 1; Length 661;
Best Local Similarity 26.4%; Pred. No. 3.7e-16;
Matches 93; Conservative 49; Mismatches 136; Indels 74; Gaps 20;

QY 13 ELITGSWS--DQTYPEGQAIYKCPGRSLNVMVCKRGSMV-----ALNPLRKQ 63
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 157 ELYNGNSTQKTFKFKVKQKYQECATGYTTAG-----KTEFEVCLTYGMSLTF--KCT 209
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 64 KRPG-----HPGDTPRGTTTLTGANVEYGVAAVYTCNEGYYLLGHEINRECDT 113
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 210 KIKCSLRLIENGYPHPK-----QYEEGDVVOEFCHENYYLSGS--DLIQCN 257
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 114 DGMWTDPICE--VYKCLPYTAPENGKIVSSAMEDREYHFOAVRFVNSGKTEGDE 171
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 258 FGWYESPVCERRRRRCRPPPLINSKIQTHS----TYYRGEIYHICELNFEIHGSAE 313
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 172 MHCSDDGFWSKEPKVCE---ISKSPDVI--NGSPISOKITIKENERFQYCKNGYEYS 226
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 314 IRC-EDGKWT-EPKPCIGOEKVAEEPPFTEINGAANLHRSKIYYNGDKVYATKCKSGYLLH 371
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 227 ERGDAVCHESGMRPLPSCDE--KSCDN--PIYIPNGYSP-LRKHTGSEIITYQCNNGYTP 282
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 372 GSNELTTCRKGMWTLPEECVENNENCKHPVVMNGAVADGLIASYATGSSVEYRC-NEYYL 430
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 283 AIRGNTAKCTSTGWIIPARCTLKPC---DYPD-----IKHGGL 317
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 431 LRGSKISRCEGQKMSPPVC-LEPCTVAVDVMNRRNINIMKKMYEKGVLHGDL 481
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
VCP_VACCV STANDARD; PRT; 263 AA.
ID VCP_VACCV
AC P10998;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE COMPLEMENT CONTROL PROTEIN PRECURSOR (VCP) (SECRETORY PROTEIN 35)
   (PROTEIN C3) (28 KDA PROTEIN).
OS Vaccinia virus (strain WR), and Vaccinia virus (strain Copenhagen).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC orthopoxvirus.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 20-37.
RC STRAIN-WR;
RX MEDLINE; 88318974.
RA Kotwal G.J., Moss B.;
RT "Vaccinia virus encodes a secretory polypeptide structurally related
   to complement control proteins."
RL Nature 335:176-178(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-WR;
RX MEDLINE; 89073756.
RA Kotwal G.J., Moss B.;
RT "Analysis of a large cluster of nonessential genes deleted from a
   vaccinia virus terminal transposition mutant."
RL Virology 167:524-537(1988).
RN [3]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-COPENHAGEN;
RA MEDLINE: 91021027.
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paolletti E.;
RA "The complete DNA sequence of vaccinia virus.";
RA Virology 179:247-266(1990).
RL [4]
RP COMPLETE GENOME.
RC STRAIN-COPENHAGEN;
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paolletti E.;
RA Virology 179:517-563(1990).
RL [5]
RP FUNCTION.
RX MEDLINE: 92115714.
RA Isaacs S.N., Kotwal G.J., Moss B.;
RT "Vaccinia virus complement-control protein prevents
RT antibody-dependent complement-enhanced neutralization of infectivity
RT and contributes to virulence.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:628-632(1992).
RN [6]
RP STRUCTURE BY NMR OF 146-263.
RX MEDLINE: 97446168.
RA Wiles A.P., Shaw G., Bright J., Percel A., Campbell I.D.,
RA Barlow P.N.;
RT "NMR studies of a viral protein that mimics the regulators of
RT complement activation.";
RL J. Mol. Biol. 272:253-265(1997).
CC -1- FUNCTION: SERVES TO PROTECT THE VIRUS AGAINST COMPLEMENT ATTACK BY
CC INHIBITING BOTH CLASSICAL AND ALTERNATIVE PATHWAYS OF COMPLEMENT
CC ACTIVATION. BINDS C3B AND C4B.
CC -1- SIMILARITY: BELONGS TO THE SUPERFAMILY OF THE REGULATORS OF
CC COMPLEMENT ACTIVATION (RCA).
CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X13166; CAA31564.1; -
DR EMBL: M2812; AAA69605.1; -
DR EMBL: A31005; WMVZSP.
DR PIR: A31005; WMVZSP.
DR PDB: 1VVC; 03-DEC-97.
DR PDB: 1VVD; 03-DEC-97.
DR PDB: 1VVE; 03-DEC-97.
DR INTERPRO: IPR000436; -.
DR PFAM: PF00084; Sush1; 4.
DR Signal: Repeat; Sush1; 3D-structure.
FT CHAIN 1 19
FT DOMAIN 20 263 COMPLEMENT CONTROL PROTEIN.
FT REPEAT 20 82 4 X SUSHI (SCR) REPEATS.
FT REPEAT 85 144 SUSHI 1.
FT REPEAT 147 202 SUSHI 2.
FT REPEAT 205 262 SUSHI 3.
FT REPEAT 262 262 SUSHI 4.
FT DISULFID 21 70 BY SIMILARITY.
FT DISULFID 54 81 BY SIMILARITY.
FT DISULFID 86 126 BY SIMILARITY.
FT DISULFID 112 143 BY SIMILARITY.
FT DISULFID 148 190 BY SIMILARITY.
FT DISULFID 176 201 BY SIMILARITY.
FT DISULFID 206 248 BY SIMILARITY.
FT DISULFID 234 261 BY SIMILARITY.
SQ SEQUENCE 263 AA; 28629 MW; EA322CC9A6EF997 CRC64;

```

Query Match

16.1%; Score 301.5; DB 1; Length 263;

```

Best Local Similarity 33.3%; Pred. No. 1.7e-16;
Matches 78; Conservative 29; Mismatches 106; Indels 21; Gaps 12;

OY 20 SDQTYPEGTQAIKCRPGYR--SLGNVIMYCKRKEGVALNPLRKCKQKRCPCGHPGPTFCT 77
DB 40 AANANIGDITIELICPGYRKOKMGPYAKCTGTTLFN---QC1KRCRCPRIDNDQ 96
OY 78 FTLTGNNPFEYGVKAVYCNCEGOLLGEL-NYRECDTDS---WTNDIPICEVKKCLPVA 133
DB 97 IDIGG---VDFSSITYSNCSGYHLGSKSYCELGSTGSMWNPAPICESVKKQSPPS 153
OY 134 PENGKIVSSAMPDREYHFGQAVRFVNCNGYKIEGDEHMCSDDGFSWKEPKVEISCK 193
DB 154 ISNGR--HNGYED--FYTGSGVYITSCNCGYSLIGNSVGLCS--GGKMS--DPTCQIVKCP 207
OY 194 SPDVINGSPISQ-KIITYKENERFYKCNNGYEXSERGDVCTESG-WRP-LDSC 244
DB 208 HPTISNGYLSGCFKRSYSYNDVDFCKYKGLSGSSSTCSFGNTWKPELPKC 261

RESULT 12
ID F13B_MOUSE STANDARD; PRT; 668 AA.
AC 007968;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE COAGULATION FACTOR XIII B CHAIN PRECURSOR (BC 2.3.2.13) (PROTEIN-
DE GLUTAMINE GAMMA-GLUTAMYLTRANSFERASE B CHAIN) (TRANSGLUTAMINASE B
DE CHAIN).
GN F13B OR F13B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN-B10.D2/OSN; TISSUE-LIVER;
RX MEDLINE: 93224141.
RA Nonaka M., Matsuda Y., Shirosaki T., Moriwaki K., Nonaka M.,
RA Natsunne-Sakai S.;
RT "Molecular cloning of the b subunit of mouse coagulation factor XIII
RT and assignment of the gene to chromosome 1: close evolutionary
RT relationship to complement factor H.";
RL Genomics 15:535-542(1993).
CC -1- FUNCTION: THE B CHAIN OF FACTOR XIII IS NOT CATALYTICALLY ACTIVE,
CC BUT IS THOUGHT TO STABILIZE THE A SUBUNIT AND REGULATE THE RATE
CC OF TRANSGLUTAMINASE FORMATION BY THROMBIN.
CC -1- SUBUNIT: TETRAMER OF TWO A CHAINS AND TWO B CHAINS.
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN LIVER AND KIDNEY.
CC -1- SIMILARITY: CONTAINS 10 SUSHI (SCR) REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D10071; BAA00963.1; -
DR PIR: A46013; A46013.
DR HSSP: P10998; 1VVC.
DR MGD: MGT:88379; F13B.
DR INTERPRO: IPR000436; -.
DR PFAM: PF00084; Sush1; 8.
DR Transferase; Plasma; blood coagulation; Repeat; Glycoprotein; Signal;
KW Sush1.
FT CHAIN 1 20 BY SIMILARITY.
FT DOMAIN 21 668 COAGULATION FACTOR XIII B CHAIN.
FT REPEAT 24 647 SUSHI 1.
FT REPEAT 90 147 SUSHI 2.
FT REPEAT 152 209 SUSHI 3.

```

FT	REPEAT	212	268	SUSHI 4.
FT	REPEAT	273	328	SUSHI 5.
FT	REPEAT	335	390	SUSHI 6.
FT	REPEAT	395	451	SUSHI 7.
FT	REPEAT	453	516	SUSHI 8.
FT	REPEAT	523	579	SUSHI 9.
FT	REPEAT	581	647	SUSHI 10.
FT	DISULFID	25	76	BY SIMILARITY.
FT	DISULFID	59	87	BY SIMILARITY.
FT	DISULFID	91	135	BY SIMILARITY.
FT	DISULFID	118	146	BY SIMILARITY.
FT	DISULFID	153	197	BY SIMILARITY.
FT	DISULFID	180	208	BY SIMILARITY.
FT	DISULFID	213	255	BY SIMILARITY.
FT	DISULFID	241	267	BY SIMILARITY.
FT	DISULFID	274	316	BY SIMILARITY.
FT	DISULFID	302	327	BY SIMILARITY.
FT	DISULFID	336	378	BY SIMILARITY.
FT	DISULFID	364	389	BY SIMILARITY.
FT	DISULFID	396	439	BY SIMILARITY.
FT	DISULFID	425	450	BY SIMILARITY.
FT	DISULFID	454	505	BY SIMILARITY.
FT	DISULFID	486	515	BY SIMILARITY.
FT	DISULFID	524	567	BY SIMILARITY.
FT	DISULFID	553	578	BY SIMILARITY.
FT	DISULFID	582	636	BY SIMILARITY.
FT	DISULFID	616	646	BY SIMILARITY.
FT	CARBONYD	162	162	N-LINKED (GLCNAC..)
FT	CARBONYD	162	162	(POTENTIAL).
FT	CARBONYD	162	162	(POTENTIAL).
SO	SEQUENCE	668 AA;	76078 MW;	BOBCE00A9E53FA6 CRC64;

	FT	CARBOHYD	665	665	N-LINKED (GLCNAC. . .) (POTENTIAL).
	FT	CARBOHYD	716	716	N-LINKED (GLCNAC. . .) (POTENTIAL).
	FT	CARBOHYD	723	723	N-LINKED (GLCNAC. . .) (POTENTIAL).
	FT	CARBOHYD	741	741	N-LINKED (GLCNAC. . .) (POTENTIAL).
	FT	LIPID	807	807	PALMITATE.
	FT	SITE	818	821	ENDOCYTOSIS SIGNAL (PROBABLE).
	FT	VARIANT	331	331	S -> N.
	FT	VARIANT	603	603	N -> D.
	FT	VARIANT	640	640	/FTid-VAR_004193.
	FT	VARIANT	756	756	L -> V.
	FT	VARIANT	756	756	/FTid-VAR_004194.
	FT	VARIANT	756	756	T -> P (REDUCED FREQUENCY IN PATIENTS WITH MYOCARDIAL INFARCTION).
	FT	VARIANT	756	756	/FTid-VAR_004195.
	SO	SEQUENCE	830 AA;	90844 MW;	FBC407BA2579FEBB CRC64;
		Query Match	15.2%	Score 286;	DB 1; Length 830;
		Best Local Similarity	29.6%;	Pred. No. 9, 6e-15;	
		Matches 87;	Conservative 36;	Mismatches 135;	Indels 34; Gaps
Qy		27	GTOAIYKRCPRYSRLSGNIYWC-RKEGVALNPLRKQCKRPGHPTP-FGTFTLTGG-	83	
Db		347	GSSCKFECPGYRVRRGLDMLRCIDSQHMSA--PLPTCEAISL-EPLSPVHGSDCPSL	4030	
Qy		84	NVEFGKAAVTTCNEGVYLLGEINRECDPDG-WTNDIPICEEYVKCLPYAPENGIYVS	1424	
Db		404	RAFQDTNCSFCAGEFLMKRADIVR-CDNIGQMTAPRPVCQAQQODLPVPNEAR--N	466	
Qy		143	AMEPDREYHFQOAVRFVNCNSGYKIIEGDDEMHSCDDGFWSKEPKVCLEISK--SPDVIN	1995	
Db		461	CSHPGARRYSVCSFTCNBEGLLVGASVLOCLATGMNNNSVPPPCQAIPTPLISPQ--N	518	
Qy		200	GS----PISQIKIYKENERFYQKNMGYEYS--ERGDAVCTESG-WRPL-PSEEEKSCD	255	
Db		519	GMTVCQVLGSS-SYKST--COFLCDEGYSLSGEERLD--CTRSGRWTDSPMCEAIKP	573	
Qy		251	NPYPNG---DYSLRIKHRTGDETTCYORRGFFPATPNAGNAKCMSG-WIIPAP	300	
Db		574	ELFAPEQSLDCSDTRGEPFNVGSTCFHCSCNNGF-KLEGPNNVECTSGRMSATP	626	
		RESULT 14			
		APOH_MOUSE	STANDARD:	PRT;	345 AA.
		AC Q01339;			
DT		01-APR-1993 (Rel. 25, Created)			
DT		01-APR-1993 (Rel. 25, Last sequence update)			
DT		01-OCT-2000 (Rel. 40, last annotation update)			
DE		BETA-2-GLYCOPROTEIN I PRECURSOR (APOLIPOPROTEIN H) (APO-H) (B2GP1)			
GN		(BETA(2))GP1 (ACTIVATED PROTEIN C-BINDING PROTEIN) (APC INHIBITOR).			
DN		APOH OR B2GP1.			
OQ		Mus musculus (Mouse).			
OC		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
RN		[1]			
RP		SEQUENCE FROM N.A.			
RX		MEDLINE: 92372000.			
RA		Nonaka M., Matsuda Y., Shirosaki T., Moriwak K., Natsunne-Sakai S.;			
RT		"Molecular cloning of mouse beta 2-glycoprotein I and mapping of the			
RL		gene to chromosome 11."			
RL		Genomics 13:1082-1087(1992).			
RN		[2]			
RP		SEQUENCE FROM N.A.			
RC		STRAIN=CBA/J; TISSUE=LIVER;			
RX		MEDLINE: 94242017.			
RA		Sellar G.C., Steel D.M., Zafiropoulos A., Seery L.T.,			
RA		Whitehead A.S.;			
RT		"Characterization, expression and evolution of mouse beta 2-			
RT		glycoprotein I (apolipoprotein H).";			
RL		Biochem. Biophys. Res. Commun. 200:1521-1528(1994).			
RN		[3]			

RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/C; TISSUE=LIVER;  
 RA Kristensen T.;  
 RT "Structure of the human beta-2-glycoprotein 1 gene."  
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: BINDS TO VARIOUS KINDS OF NEGATIVELY CHARGED SUBSTANCES  
 CC SUCH AS HEPARIN, PHOSPHOLIPIDS, AND DEPRAN SULFATE. MAY PREVENT  
 CC ACTIVATION OF THE INTRINSIC BLOOD COAGULATION CASCADE BY BINDING  
 CC TO PHOSPHOLIPIDS ON THE SURFACE OF DAMAGED CELLS.  
 CC -1- TISSUE SPECIFICITY: PLASMA.  
 CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: D10056; BAA00945.1; -  
 DR EMBL: S70439; AAB30789.1; -  
 DR EMBL: Y11356; CAA72190.1; -  
 DR PIR: A43286; NBMS.  
 DR HSSP: P10998; IYVC.  
 DR MGD: MGI:88058; APOH.  
 DR INTERPRO: IPR000436; -  
 DR PFAM: PF00084; sushi; 4  
 KW Heparin-binding; Glycoprotein; Plasma; Repeat; Sushi; Signal.  
 FT SIGNAL 1 19  
 FT CHAIN 20 345  
 FT DOMAIN 22 261  
 FT REPEAT 22 80  
 FT REPEAT 83 138  
 FT REPEAT 141 201  
 FT REPEAT 204 261  
 FT DOMAIN 263 345  
 FT DISULFID 23 66  
 FT DISULFID 51 79  
 FT DISULFID 84 124  
 FT DISULFID 110 137  
 FT DISULFID 142 188  
 FT DISULFID 174 200  
 FT DISULFID 205 248  
 FT DISULFID 234 260  
 FT DISULFID 264 315  
 FT DISULFID 300 325  
 FT DISULFID 307 345  
 FT CARBOHYD 162 162  
 FT CARBOHYD 183 183  
 FT CARBOHYD 193 193  
 FT CONFLICT 252 252  
 SQ SEQUENCE 345 AA; 38619 MW; C83F8A6EBD51C940 CMC64;  
 Query Match 14.98; Score 279.5; DB 1; Length 345;  
 Best Local Similarity 27.78; Pred. No. 1.1e-14;  
 Matches 90; Conservative 44; Mismatches 132; Indels 59; Gaps 19;  
 QY 23 TYPGTAIYCRPGYRSLGNV-IMVC-RKGEWALNLRCKQKRPCHPDTPRGTTL 80  
 Db 40 SYDGEQIVISCKRGYISRGGMKRFCTCLTGM-PINTLR-CVPPVCPFAGLENGIVRY 97  
 QY 81 TGNVFEYGVKAVYTCNEGVLCEINRECDTGC-WTNDIPICEVYVCLPVTAPENGKI 139  
 Db 98 TS--FEYPKNISFACNPGFPLNG-TSSSKCTEGSKMSPDIPACARITCPPPPPKFALL 153  
 QY 140 VSSAMEDEDRHYEQOARFPCVNSCYKIEGDEMAHCSDDGFSKSKPKCVETISCK-SPDYI 198  
 Db 154 KDRPSAGNNSLVQDYTVVFCPLPHFAMIIGNDTVMCTEGNMWTR-LPBCLEVKCPFPPE 212  
 QY 199 NG---SPISOKIYKENEROYKCMNGEYSEKGDVCTESG-WRLPLSCSEKSCNDPIYI 254

Db 213 NGVYVPAKPEVLLYKDKATE--GCHETVKLDGPEAECECTGTGWSFLPTCRE-SCKLP-V 268  
 QY 255 PNGD--YSPRIKHT-----GDEITYQCNGNGFFPARCNTAKTST-----GWI 297  
 Db 269 KKATVLYQGRKRVKIQEFGKMMHGDITHYCKN-----KCKSYTVEAHCRDGTI 320  
 QY 298 PADRC-----TLKPC 307  
 Db 321 EIPSCFEHSSLAFWKUDASELTPC 345  
 RESULT 15  
 ID LEM2\_HUMAN  
 ID LEM2\_HUMAN STANDARD; PRT; 610 AA.  
 AC P16581; P16111;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE E-SELECTIN PRECURSOR (ENDOTHELIAL LEUKOCYTE ADHESION MOLECULE 1)  
 DE (ELAM-1) (LEUKOCYTE-ENDOTHELIAL CELL ADHESION MOLECULE 2) (LECAM2)  
 DE (CD62E).  
 OS Homo sapiens (Human).  
 GN Sele or ELAM1.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 90175359.  
 RA Hession C., Osborn L., Goff D., Chi-Rosso G., Vassallo C.,  
 RA Pasek M., Pittack C., Tizard R., Goetz S., McCarthy K., Hopple S.,  
 RA Lobb R.,  
 RT "Endothelial leukocyte adhesion molecule 1: direct expression cloning  
 RT and functional interactions."  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:1673-1677(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 89162047.  
 RA Bevilacqua M.P., Stengelin S., Gimbrone M.A. Jr., Seed B.;  
 RT "Endothelial leukocyte adhesion molecule 1: an inducible receptor for  
 RT neutrophils related to complement regulatory proteins and lectins."  
 RL Science 243:1160-1165(1989).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 91115870.  
 RA Collins T., Williams A., Johnston G.I., Kim J., Eddy R., Shows T.,  
 RA Gimbrone M.A. Jr., Bevilacqua M.P.;  
 RT "Structure and chromosomal location of the gene for endothelial-  
 RT leukocyte adhesion molecule 1."  
 RL J. Biol. Chem. 266:2466-2473(1991).  
 RN [4]  
 RP LIGAND.  
 RX MEDLINE: 91068005.  
 RA Phillips M.L., Nudelman E., Gaeta F.C., Perez M., Singhal A.K.,  
 RA Hakomori S., Paulson J.C.;  
 RT "ELAM-1 mediates cell adhesion by recognition of a carbohydrate  
 RT ligand, sialyl-Lex."  
 RL Science 250:1130-1132(1990).  
 RN [5]  
 RP 3D-STRUCTURE MODELLING OF LECTIN DOMAIN.  
 RX MEDLINE: 93202275.  
 RA Mills A.;  
 RT "Modelling the carbohydrate recognition domain of human E-selectin."  
 RL FEBS Lett. 319:5-11(1993).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 22-178.  
 RX MEDLINE: 94150646.  
 RA Graves B.J., Crowther R.L., Chandran C., Rumberger J.M., Li S.,  
 RA Huang K.-S., Presky D.H., Familletti P.C., Wollittky B.A., Burns D.K.;  
 RT "Insight into E-selectin/ligand interaction from the crystal  
 RT structure and mutagenesis of the Iec/Egf domains."  
 RL Nature 367:532-538(1994).  
 RN [7]  
 RP VARIANT ARG-149.

Query Match	Best Local Similarity	Score 279;	DB 1;	Length 610;
Matches 82;	Conservative 44;	Mismatches 137;	Indels 52;	Gaps 19;
QY 23 TYPEGTQATIKCRPGYRSLGNVIMWC-RKGEVVALNPLRK---CQ--KRP-----CGHP 70	14.9%;	Score 279;	DB 1;	Length 610;
DB 261 SFWMTTCTCFDECEGFEMLGAOSLIQTSISGNDMNEKPTCKAVYICRAVRQPONGSVYRCSH- 319	26.0%;	Pred. No. 2.4e-14;		
QY 71 GDRPFTFTLLTGNAVREYGVKANYTCNDEGYTLGLGELNRECDPDG-WINDIPICEVYVKLT 129	82;	Conservative 44;	Mismatches 137;	Indels 52;
DB 320 --SPACEF-----FKSSCNFTCEBEGMLOGPAOV-ECTTGGMTQOIIYVCEAFQCT 368	26.0%;	Pred. No. 2.4e-14;		
QY 130 PYTAPENGRIVSSAMPDEHYHFGQAVRVCVNSGYKITEGDEMHCSDDGFWSKEKPCYE 189	82;	Conservative 44;	Mismatches 137;	Indels 52;
DB 369 ALSNPERG-YMNCILPASAASFYRGSCPECEGDFYIKGSKRLQCGPTGEMDKPTCEA 427	26.0%;	Pred. No. 2.4e-14;		
QY 190 ISCKSPDVIN-----GSPISKIIYKENRQYOKCMKGYEYSEKGDVACTESSG-W 238	82;	Conservative 44;	Mismatches 137;	Indels 52;
DB 428 VRC--DAYHQPKGLVRCAHSPIGE-FYTKSSCAF--SCEBEFEHLHGSTOLECTYSOGGM 481	26.0%;	Pred. No. 2.4e-14;		
QY 239 -RLPSCCEESKNPFIIPNGDYSPLRIKIHRTGDEILITYOCNNGFYPATRGNTAK-CTSGT- 295	82;	Conservative 44;	Mismatches 137;	Indels 52;
DB 482 TEEVPSCCQVYKCSLAVPGKINMSCGGEPIYGVICAFACPEGM--TLNLSAANTGCATGCH 539	26.0%;	Pred. No. 2.4e-14;		
QY 296 WIPAPRCTLKPCDYP 310	82;	Conservative 44;	Mismatches 137;	Indels 52;
DB 540 W-----SGLLPTCEAP 550	26.0%;	Pred. No. 2.4e-14;		

Tue Nov 21 16:57:30 2000

us-09-316-163-11.rsp

Page 18

Search completed: November 21, 2000, 16:49:24  
Job time: 327 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2000, 16:48:51 ; Search time 70.54 Seconds  
(without alignments)  
435.499 Million cell updates/sec

Title: US-09-316-163-11

Perfect score: 1876

Sequence: 1 EDCNLEPRRNTTEITGSMS.....PDIKHGLYHEMRPRFPV 329

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 297973 seqs, 93374136 residues

Total number of hits satisfying chosen parameters: 297973

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPTREMBL\_14:\*  
2: sp.archaea:\*  
3: sp.bacteria:\*  
4: sp.fungi:\*  
5: sp.human:\*  
6: sp.invertebrate:\*  
7: sp.mammal:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp.protist:\*  
12: sp.virus:\*  
13: sp.vertebrate:\*  
14: sp.unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1876	100.0	449	4	014570
2	903	48.1	669	6	028085
3	486	25.9	1053	13	091275
4	377	20.1	522	6	028769
5	353	18.8	679	11	029524
6	351	18.7	1911	6	029528
7	350.5	18.7	559	4	090072
8	350.5	18.7	2039	4	016745
9	350.5	18.7	2489	4	016744
10	349.5	18.6	645	12	09WPU2
11	348	18.6	2014	6	029530
12	342.5	18.3	661	6	029531
13	339	18.1	522	6	028797
14	337.5	18.0	560	5	022328
15	335	17.9	360	12	09WU08
16	333.5	17.8	315	6	028770
17	328	17.5	550	12	P88903
18	328	17.5	550	12	040912
19	323	17.2	497	11	063612

20	323	17.2	559	11	063135	063135 rattus norv
21	317	16.9	452	11	061407	061407 mus musculu
22	310	16.5	255	12	P87616	P87616 compox viru
23	308.5	16.4	1032	4	013866	013866 homo sapien
24	306.5	16.3	417	11	035520	035520 rattus norv
25	306	16.3	303	11	061405	061405 mus musculu
26	304	16.2	483	11	064735	064735 mus musculu
27	302.5	16.1	657	4	014006	014006 homo sapien
28	302	16.1	533	11	008569	008569 cavia porce
29	298	15.9	1124	5	09VIR4	09VIR4 drosophila
30	294	15.7	1652	5	09VIR9	09VIR9 drosophila
31	293	15.6	363	6	002839	002839 sus scrofa
32	292.5	15.6	1045	6	046545	046545 ovis aries
33	290	15.5	1087	4	014212	014212 homo sapien
34	285	15.2	579	11	060736	060736 mus musculu
35	284.5	15.2	263	12	098659	098659 variola vir
36	284.5	15.2	263	12	007033	007033 variola vir
37	284	15.1	740	4	095508	095508 homo sapien
38	283.5	15.1	263	12	089076	089076 variola vir
39	283.5	15.1	974	5	P91658	P91658 drosophila
40	280.5	15.0	263	12	089061	089061 variola vir
41	278	14.8	354	5	090611	090611 drosophila
42	274.5	14.6	349	4	015429	015429 homo sapien
43	272	14.5	285	6	019126	019126 macaca fasc
44	272	14.5	285	6	019127	019127 macaca mula
45	270.5	14.4	369	6	P79138	P79138 cercopithec

## ALIGNMENTS

RESULT 1  
ID 014570 PRELIMINARY: PRT: 449 AA.  
AC 014570: P78435;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
DE COMPLEMENT FACTOR H PRECURSOR.  
GN HF OR CFH.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
[1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 88134059.  
RA Ripoche J., Day A.J., Harris T.J.R., Sim R.B.;  
RT "The complete amino acid sequence of human complement factor H";  
RL Biochem. J. 249:593-602(1988).  
[2]  
RP SEQUENCE OF 226-449 FROM N.A.  
RX MEDLINE: 86169701.  
RA Kristensen T., Wetsel R.A., Tack B.F.;  
RT "Structural analysis of human complement protein H: homology with C4b  
binding protein, beta 2-glycoprotein I, and the Ba fragment of B2";  
RL J. Immunol. 136:3407-3411(1986).  
[3]  
RP SEQUENCE OF 1-19 FROM N.A.  
RX Vix D.P., Williams S.A.;  
RT Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.  
[4]  
RP SEQUENCE OF 1-9 FROM N.A.  
RX Dominguez O.;  
RT Thesis (1993), Immunologia, Hospital Trias I Pujol, Spain.  
EMBL: X07523; CAAS0403.1; -;  
DR EMBL: M12383; AAS2013.1; -;  
DR EMBL: U56979; AAB01987.1; -;  
DR EMBL: Z29655; CAA82763.1; -;  
DR HSSP: P10998; IYVC.  
DR INTERPRO: IPR000436; -;  
DR PFWA: PF000084; sushi; 7.  
KW SIGNAL.  
FT SIGNAL 1 18 POTENTIAL.

FT CHAIN 19 449 POTENTIAL.  
SQ SEQUENCE 449 AA: 51033 MW: C2AAD4F155343E3 CRC64;

Query Match 100.0%; Score 1876; DB 4; Length 449;  
Best Local Similarity 100.0%; Pred. No. 6.8e-161;  
Matches 329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDCNELPPRRNTILGMSWSDQTYTPECTQAIYKCRPGYBSLGNVIMVCRGGEVVALPLR 60  
DB 19 EDCNELPPRRNTILGMSWSDQTYTPECTQAIYKCRPGYBSLGNVIMVCRGGEVVALPLR 78  
QY 61 KCGKRCGHHGDPFGFTLTGNGVFEYGVKAYTTCNEGQQLGELINRECDTDGWTNDI 120  
DB 79 KCGKRCGHHGDPFGFTLTGNGVFEYGVKAYTTCNEGQQLGELINRECDTDGWTNDI 138  
QY 121 PICEVVKCLPVTAPENGKIVSSAMEPDRREYHFGQAVRFVNSGKIEGDEMHCSDDGFV 180  
DB 139 PICEVVKCLPVTAPENGKIVSSAMEPDRREYHFGQAVRFVNSGKIEGDEMHCSDDGFV 198  
QY 181 SKRKPCKVEISCKSPDVINGSPISOKITIKENRFPQYKCMGYEYSEKDAVCTESGWRP 240  
DB 199 SKRKPCKVEISCKSPDVINGSPISOKITIKENRFPQYKCMGYEYSEKDAVCTESGWRP 258  
QY 241 LPSCSEKSCDNPYIPNGDYSPLRIKRTGDEITYOCNNGFYPATRGNTAKCTSTGWTIPAP 300  
DB 259 LPSCSEKSCDNPYIPNGDYSPLRIKRTGDEITYOCNNGFYPATRGNTAKCTSTGWTIPAP 318  
QY 301 RCTLKPCDYPDIKHGGLYHENMRPYPV 329  
DB 319 RCTLKPCDYPDIKHGGLYHENMRPYPV 347  
RESULT 2  
ID Q28085 PRELIMINARY; PRT; 669 AA.  
AC Q28085;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)  
DT 01-MAY-2000 (Tremblrel. 13, last annotation update)  
DE CCP MODULES 3-12, WITH PARTS OF CCP 2 AND 13 (FRAGMENT).  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER;  
RX MEDLINE: 96202005.  
RA Soames C.J., Day A.J., Sim R.B.;  
RT "Prediction from sequence comparisons of residues of factor H involved  
in the interaction with complement component C3b.";  
RL Biochem. J. 315:523-531(1996).  
DR EMBL: X98697; CA67257.1; -.  
DR HSSP: P10998; IYV.  
DR INTERPRO: IPR000436; -.  
DR PFAM: PF00084; sush1, 11.  
FT NON\_TER 1 1  
FT TER 669 669  
SQ SEQUENCE 669 AA: 75683 MW: DDD9DB30E747AC2 CRC64;

Query Match 48.1%; Score 903; DB 6; Length 669;  
Best Local Similarity 60.9%; Pred. No. 3.8e-73;  
Matches 154; Conservative 33; Mismatches 64; Indels 2; Gaps 1;

QY 76 GTFTLTGNGVFEYGVKAYTTCNEGQQLGELINRECDTDGWTNDIPICEVVKCLPVTAP 135  
DB 3 GSPHAEENOFEGYGAQVYVTCDEGYQWVGMNFRCDTNGMTNDIPICEVVKCLPVTAP 62  
QY 136 NGKITSAMFDPREYHFGQAVRFVNSGKIEGDEMHCSDDGFMSKPKVEISCKRP 195  
DB 63 NGKITSAMFDPREYHFGQAVRFVNSGKIEGDEMHCSDDGFMSKPKVEISCKRP 122

QY 196 DVINGSPISOKIYKENERFQYKCMNGEYSEKDAVCTESGWRPPLPSCSEKSCDNPYIP 255  
DB 123 VIINGQAVLPKATYKONERQYKCAAGFEGQREDYVCTSGMPPATCTGTEITDDPPRIP 182  
QY 256 NGDYSPLRIHRTGDEITYOCNNGFYPATRGNTAKCTSTGWTIPAPRCTLKPCDYPDIKHG 315  
DB 183 NGYRPELSKXRGQDKITTECKKGFPEIRGTATCTRDGWPVPRCAMKPCSYPIKHG 242  
QY 316 GLYHENMRPYPV 328  
DB 243 RLYYS--YRGYFP 253

RESULT 3  
ID Q91275 PRELIMINARY; PRT; 1053 AA.  
AC Q91275;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)  
DT 01-MAY-2000 (Tremblrel. 13, last annotation update)  
DE COMPLEMENT REGULATORY PLASMA PROTEIN.  
OS Paratubrax nebulifer (barred sand bass).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Acanthopterygii;  
OC Perciformes; Percloides; Serranidae; Paratubrax.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER;  
RX MEDLINE: 94318039.  
RA Dahmen A., Kaidoh T., Zipfel P.F., Gagli I.;  
RT "Cloning and characterization of a cDNA representing a putative  
complement-regulatory plasma protein from barred sand bass (Paratubrax  
nebulifer)."  
RL Biochem. J. 301:391-397(1994).  
DR EMBL: L21703; AAA92556.1; -.  
DR HSSP: P08603; IHP.  
DR INTERPRO: IPR000436; -.  
DR PFAM: PF00084; sush1, 16.  
SQ SEQUENCE 1053 AA: 117597 MW: F2E32C3AD76D5D3 CRC64;

Query Match 25.9%; Score 486; DB 13; Length 1053;  
Best Local Similarity 29.7%; Pred. No. 2.1e-35;  
Matches 114; Conservative 48; Mismatches 134; Indels 88; Gaps 13;

QY 21 DOTYPTGQAIYKCRPGYBSLGNVIMVCRGGEVVALPLRCKCRPGHGDTPFGFTL 80  
DB 47 EASYPGRGQVRCANVGS--GFFKLVCEGKWTMG--AKQPRSCGHHGDAQFADPHL 102  
QY 81 TGNVFEYGVKAYTTCNEGQQLGELINRECDTDGWTNDIPICEVVKCLPVTAPENGKIV 140  
DB 103 AECNDFVFGSKVYVTCQKGYQWVSRINRYRCVAEGMDGVVPCESQOCPLIHNDNVQYI 162  
QY 141 SSAMEPDRREYHFGQAVRFVNSGKIEGDEMHCSDDGFMSKPKVEISCKSPDVIN 199  
DB 163 GG-----PEATGNGVVRFSCKSKSEILDSPELYCDERGDMSGPVYKCAITCAIPDEN 218  
QY 200 GSPISOKIYKENERFQYKCMNGEYSEKDAVCTESG---WRPLPSCSEKSC----- 249  
DB 219 GNVPGAIRREKENDVLDHCBDAFKHIDR-PSICIQGIIKAEMSPPLCESIKCRTIIND 277  
QY 250 -----DNPIIP-----NGDYS----- 260  
DB 278 GTRYEPAYRNLPSPGFTLVICARTSWISTPQETSVTTCODNGEWSIRPTQOEVRCSNR 337  
QY 261 -PLRI-----KRTGDEITYOCNNGFYPATRGNT-AKCTSGWTIPAPRCTLKPCDY 309  
DB 338 REHDSMVRSMERTLTDNRTYCKRG-YKRTGVTYATCGRNGMNPPLCEVYTKSK 396  
QY 310 PDIKHG-----GLYHENMRPYPV 326  
DB 397 ENIODAVIVGDKQIYNLNQKAIY 420

RESULT 4  
ID Q28769 PRELIMINARY; PRT; 522 AA.  
AC Q28769;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, last sequence update)  
DT 01-MAY-2000 (TREMblrel. 13, last annotation update)  
DE COMPLEMENT RECEPTOR (FRAGMENT).  
OS Papio hamadryas (Hamadryas baboon).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
OC Cercopithecinae; Papio.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BONE MARROW;  
RA Blumhagen D.J., Logar C.M., Shen X.P., Chen W.;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U77977; AAA99004.1; -.  
DR HSSP; P08603; IHTI.  
DR INTERPRO; IPR000436; -.  
DR PFAM; PF00084; sush1; 7.  
FT NON\_TER 1  
SQ SEQUENCE 522 AA; 56626 MW; 312FCBE03ADFL9DC CRC64;

Query Match 20.1%; Score 377; DB 6; Length 522;  
Best Local Similarity 30.2%; Pred. No. 5.8e-26;  
Matches 107; Conservative 55; Mismatches 142; Indels 50; Gaps 22;

QY 3 CN---ELPPRTEILLTSSWSDQTYPECTQAIYKCRPGYRSLGNVIMCRKGEWVALNPL 59  
DB 48 CNAPEQLFAPRTNLTDS---EPFVGYLYKECLPGHGKPFSLICLKNSVMTSADK- 102  
QY 60 RKCOARPCGDDPFGFTLTGNAVFEYKAVYTCNCGYOLLGEINREDTDC--- 115  
DB 103 -KCTKSKRNRPDPVNGAVHY--KDIOFGSIOYNSCKNGYLLIS--SATCTIISGNYI 158  
QY 116 WTNDIPICEVVKC-LPVTAPENGKIVSSAMEPDRYH-FGOAVRFVCSNG-----YKLE 167  
DB 159 WDNEPPICEIIPGGLPPTI-ANGDFISTSS---REYFPGSVVYTRCNLGSRRKLELV 213  
QY 168 GDEEMHCS--DD--GFWSKERKPC-VEISCKSPDYINGSPIS-QKIITYENERPOYCKM 221  
DB 214 GEPSTYCSKSDQVIGWGPAPQCIIPKCMPPNENGLVSNRSLFSLNEVEFERCOP 273  
QY 222 GYEYERGDVACTE-SGNRP-LPSCSEKSCDNPYIPNGDYSPLRIKHRTGDEITYQCRRG 279  
DB 274 GFVMGPRHVCOALNKKEPELPPSCSRVCCQPPPELHGHETSQDPSGQEVFVSCPEG 333  
QY 280 FYPATRGNTAKCTSTG-WIP-APRCTLKPCD--YPDIKHGGLYHENMRPPYFV 329  
DB 334 -YDLGASLSHCTPGQDWNPEAPICTVSKSCDDFLGQLPHG-----RVLPPL 378

RESULT 5  
ID Q99254 PRELIMINARY; PRT; 679 AA.  
AC Q99254;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, last sequence update)  
DT 01-MAY-2000 (TREMblrel. 13, last annotation update)  
DE COMPLEMENT RECEPTOR TYPE 2 (CR2) (FRAGMENT).  
GN CR2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BALB/C;  
RX MEDLINE; 90229754.  
RA Kurtz C.B., O'Toole E., Christensen S.M., Weis J.H.;

RT "The murine complement receptor gene family. IV. Alternative splicing  
RT of Cr2 gene transcripts predicts two distinct gene products that share  
RT homologous domains with both human CR2 and CR1."  
RL J. Immunol. 144:3581-3591(1990).  
RN [2]  
RP SEQUENCE OF 21-367 FROM N.A.  
RX MEDLINE; 95105691.  
RA Kim Y.U., Kinoshita T., Molina H., Hourcade D., Sessa T., Wagner L.M.,  
RA Holter V.M.;  
RT "Mouse complement regulatory protein Cr2/p65 uses the specific  
RT mechanisms of both human decay-accelerating factor and membrane  
RT cofactor protein."  
RL J. Exp. Med. 181:151-159(1995).  
DR EMBL; U17128; AAA78271.1; -.  
DR EMBL; U17123; AAA78271.1; JOINED.  
DR EMBL; U17124; AAA78271.1; JOINED.  
DR EMBL; U17125; AAA78271.1; JOINED.  
DR EMBL; U17126; AAA78271.1; JOINED.  
DR EMBL; U17127; AAA78271.1; JOINED.  
DR EMBL; M36470; AAA37449.1; -.  
DR HSSP; P10998; IYVC.  
DR MGD; MGI:88489; Cr2.  
DR INTERPRO; IPR000436; -.  
DR PFAM; PF00084; sush1; 10.  
FT NON\_TER 679  
SQ SEQUENCE 679 AA; 74916 MW; 52FC00FDCED20CDC CRC64;

Query Match 18.8%; Score 353; DB 11; Length 679;  
Best Local Similarity 28.9%; Pred. No. 1.1e-23;  
Matches 93; Conservative 61; Mismatches 136; Indels 32; Gaps 17;

QY 20 SPQT-YPECTQAIYKCRPGYRSLGNVIMCRKGEWVALNPLRKCOARPCGHPGDPFGF 78  
DB 37 SDKSEFALGTTEYKCRPFYFRKPSFIITCIETSKW--SDAQOFCKRRKPCMNQEPHGSV 94  
QY 79 TLGNAVFEYKAVYTCNCGYOLLGEINREDTDC---DPDGMNDIPICEVVKCLPVTAP 134  
DB 95 HINTG--IEFGSTIYSCNQGRLIJD--SATCTIYSDNTVMNDMPDLCESTPCSEPPAI 151  
QY 135 ENKIVSSAMEPDRYHFGQAVRFVCSNG-----YKIGDEEMHCSDD---GFWSKER 184  
DB 152 SNGDEFSSSRD---SFFYGMVYTYCHTGKNREKFLDLVGKSIYCTSDNDNGINNSP 208  
QY 185 PKCV-EISCKSPDYINGSPISQ-KIITYENERPOYCKMNGYEYSEGDVAC-TESGWR-P 240  
DB 209 POCIPRVKCPMEIENGELVSGFKHSFPLNDYVIFKCKSGFTYKGSRIAMCQPNKWSPP 268  
QY 241 LPSCSEKSCDNPYIPNGDYSPLRIKHRTGDEITYQCRRGFPATRGNTAKCTSTG-WI-P 298  
DB 269 LPTCFMGCLPPONIIHGDYNNKDKDEFVSQGVKSYTCNPG-YTLIGNLVECTSLGTSMT 327  
QY 299 APRCTLKPCD-YPD-IKHGGLY 318  
DB 328 VPTCEVKSQDAIPIHNLHGRVF 349

RESULT 6  
ID Q29528 PRELIMINARY; PRT; 1911 AA.  
AC Q29528;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, last sequence update)  
DT 01-MAY-2000 (TREMblrel. 13, last annotation update)  
DE COMPLEMENT RECEPTOR 1 (FRAGMENT).  
GN CR1.  
OS Papio hamadryas (Hamadryas baboon).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
OC Cercopithecinae; Papio.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Clemenza L., Subramanian B.V., Nickells M.W., Hourcade D.E.;

RA Atkinson J.P.;  
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: L39791; AAA62170.1; -.  
 DR HSSP: P08603; IHCC.  
 DR INTERPRO: IPR000436; -.  
 DR INTERPRO: IPR000834; -.  
 DR INTERPRO: IPR001424; -.  
 DR PIRAM: PFO0084; sushi; 29.  
 DR PROSITE: PS00087; SOD\_CU\_ZN\_1; UNKNOWN\_1.  
 DR PROSITE: PS00133; CARBOXYPEPT\_ZN\_2; UNKNOWN\_4.  
 FT NON\_TER 1911 1911  
 SQ SEQUENCE 1911 AA: 210173 MW: 535AADD0EAF521D CRC64:

Query Match 18.7%; Score 351; DB 6; Length 1911;  
 Best Local Similarity 29.0%; Pred. No. 6.1e-23;  
 Matches 93; Conservative 57; Mismatches 135; Indels 36; Gaps 21;

QY 20 SDQTYEGTQAIKCRGYSLSGNVIMV-CRKG-EMVALNPLKCKQKRCGPHGDTTFFGT 77  
 DB 1299 NDEFPYGTSLNECHPGY--FGMFISLLENLWMSVSD--NCRKSCGTPE-PFNG 1353  
 QY 78 FTLLGNGVEYGVKAVYTCNEGYOLLGEINVRCDTGD---WTNDIPICEVVKLPVTA 133  
 DB 1354 MVAHINTDT-QFGSTVNSCNEGFRLLISPS-TTCIVSGNNVMTDKAPICEITISCKRPPT 1411  
 QY 134 PENGKIVSSAMEDREYHFGQAVFVNCNSG-----YKIGDEEMHCS--DD--GFWSKE 183  
 DB 1412 ISNGDEYSNNR---TSFSGTAVTYOCHTGPDEOLFELGERSIYCTSKDDQGAWSSP 1468  
 QY 184 KPKCVELIS-CKSPDYINGSPI-SOKIITYKENRFOYKCNNGYSESGDAVC--TESGMRP 240  
 DB 1469 PPRCISTNCTAEVANGIVPQNRFSFLNELYFRFCQPEFVWGSHTYOCOTNNRNGP 1528  
 QY 241 -LPSCEKSCDNPIYPNGDYSPI-RIKHRTGDEITYQCNGFYPAIRGNTAKTSTG-WT 297  
 DB 1529 KLPHCGRVCPPEILLHGETTSPHOKFSPQGVFVSCBPG-YDLRGAASLHCTPGQDWS 1587  
 QY 298 P-APRCTLKPCD--YPDIKHG 315  
 DB 1588 PEAPICTVKSCEDETLQQLPHG 1608

RESULT 7  
 Q9UOV2 PRELIMINARY; PRT; 559 AA.  
 AC Q9UOV2;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)  
 DE HUMAN CRI MRNA FOR C3B/C4B RECEPTOR SECRETED FORM PRECURSOR (FRAGMENT).  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 89010527.  
 RA Houchard D., Miesner D.R., Atkinson J.P., Holers V.M.;  
 RT "Identification of an alternative polyadenylation site in the human  
 RT C3b/C4b receptor (complement receptor type 1) transcriptional unit and  
 RT prediction of a secreted form of complement receptor type 1.";  
 RL J. Exp. Med. 168:1255-1270(1988).  
 DR EMBL: X14362; CAA32541.1; -.  
 DR INTERPRO: IPR000436; -.  
 DR INTERPRO: IPR000834; -.  
 DR PIRAM: PFO0084; sushi; 8.  
 DR PROSITE: PS00133; CARBOXYPEPT\_ZN\_2; UNKNOWN\_1.  
 KM Signal.  
 FT NON\_TER <1 1  
 FT SIGNAL 17 559 POTENTIAL.  
 FT CHAIN 17 559  
 SQ SEQUENCE 559 AA: 61424 MW: DBFEE965CA179D75 CRC64:

Query Match 18.7%; Score 350.5; DB 4; Length 559;  
 Best Local Similarity 29.3%; Pred. No. 1.5e-23;  
 Matches 100; Conservative 55; Mismatches 127; Indels 59; Gaps 23;

QY 3 CNE---LPRRNTIELTGSWSDQTYEGTQAIKCRGYSLSGNVIMVCRKGEWVALNPL 59  
 DB 18 CNAPEWLPFAFPTMLT---DEFEFFIGTYLNYCGRGYSGRPSITLCKNSVTGAKD- 72  
 QY 60 RKCKRPGCHGDPPTFTLLTGNVEYGVKAVYTCNEGYOLLGEINVRCDTGD---PTDG 115  
 DB 73 -KCRKSCRNPPDPVNGMVAHIK--TQFSQIKVSCTKGRLIG--SATCIITSGPTVI 128  
 QY 116 WTNDIPICEVVK-LPYTAPENGKIVSSAMEDREYHFGQAVFVNCNSG-----YKIEG 168  
 DB 129 WDNETPICDRIPGLPPTI-TNGDFISTNR---NFHGSVITYRCNPGSGGRVPELVG 184  
 QY 169 DEEMHC--SD--GFWSKEKPKC-VEISCKSPDYINGSPI-SOKIITYKENRFOYKCNMG 222  
 DB 185 EPSIYCTSNDDQVIMSGPAPOCIIPKCTPPNVENGLIVSDNLSFLNEVVEFRCPG 244  
 QY 223 YEYSEKGDVCTE-SGRP-LPSCEKSCDNPIYPN-----GDYSPLRIKHRTG 269  
 DB 245 FVAKGPRRVKQALNKKEPELPSC-SRVCP--PDVLAERTORXDNFSP-----G 294  
 QY 270 DEITYQCRNGFYPATRGNTAKTSTG-WIP-APRCTLKPCD 308  
 DB 295 QEVFYSCBPG-YDLRGAASMRCTPGQDWSPAAPICEVKSCD 334

RESULT 8  
 Q16745 PRELIMINARY; PRT; 2039 AA.  
 AC Q16745;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)  
 DE COMPLEMENT RECEPTOR 1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 94065175.  
 RA Vik D.P., Wong W.W.;  
 RT "Structure of the gene for the F allele of complement receptor type 1  
 RT and sequence of the coding region unique to the S allele.";  
 RL J. Immunol. 151:6214-6224(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Vik D.P., Wong W.W.;  
 RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: L17418; AAB60694.1; -.  
 DR EMBL: L17390; AAB60694.1; JOINED.  
 DR EMBL: L17391; AAB60694.1; JOINED.  
 DR EMBL: L17392; AAB60694.1; JOINED.  
 DR EMBL: L17393; AAB60694.1; JOINED.  
 DR EMBL: L17394; AAB60694.1; JOINED.  
 DR EMBL: L17395; AAB60694.1; JOINED.  
 DR EMBL: L17396; AAB60694.1; JOINED.  
 DR EMBL: L17397; AAB60694.1; JOINED.  
 DR EMBL: L17398; AAB60694.1; JOINED.  
 DR EMBL: L17399; AAB60694.1; JOINED.  
 DR EMBL: L17400; AAB60694.1; JOINED.  
 DR EMBL: L17401; AAB60694.1; JOINED.  
 DR EMBL: L17402; AAB60694.1; JOINED.  
 DR EMBL: L17403; AAB60694.1; JOINED.  
 DR EMBL: L17404; AAB60694.1; JOINED.  
 DR EMBL: L17405; AAB60694.1; JOINED.  
 DR EMBL: L17406; AAB60694.1; JOINED.  
 DR EMBL: L17407; AAB60694.1; JOINED.  
 DR EMBL: L17408; AAB60694.1; JOINED.  
 DR EMBL: L17409; AAB60694.1; JOINED.



```

Db      270 FVMKGPBRVRCQALMKWPELPSG-SRVCQP--PDVLHARRTRQDRDNFSP-----G 312
Qy      270 DEITYQCRRNGEYPATRGNTACTSTG-WIP-APRCTLYKPCD 308
       | : | | | : | | | : | | | : | |
Db      320 QEVFYSCEPG-YDLRGASMRCTPGGDWSPAAPCFEVXSCD 359

RESULT  10

O9WRU2  O9WRU2    PRELIMINARY;          PRt;        645 AA.
O9WRU2:
Dt      01-NOV-1999 (TREMBLrel. 12, Created)
Dt      01-NOV-1999 (TREMBLrel. 12, Last sequence update)
Dt      01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE      COMPLEMENT BINDING PROTEIN.
OS      Macaca mulatta rhadinovirus 17577.
OC      Viruses: dsDNA viruses, no RNA stage; Herpesviridae;
CC      Gammaherpesvirinae; Rhadinovirus.
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE; 99174001.
RA      Seales R.P., Bergquam E.P., Axthelm M.K., Wong S.W.;
       "Sequence and genomic analysis of a Rhesus macaque rhadinovirus with
       similarity to Kaposi's sarcoma-associated herpesvirus/human
       herpesvirus 8.";
RL      J. Virol. 73:3040-3053(1999).
DR      EMBL; AF083501; AAD21332.1; -.
DR      INTERPRO; IPR000436; -.
DR      PFMAM; PF00084; sush1; 8.
SO      SEQUENCE 645 AA; 71526 MW; 93DBDE35ABF61EB2 CRC64;
```

Query Match	Similarity	18.6%	Score 349.5	DB 12	Length 645
Best Local	Similarity	25.8%	Pred. No. 2.2e-23		
Matches	99	Conservative	45	Mismatches 164	Indels 75
					Gaps 14
QY	1	EDCNELPPRRNTEILGWSMDQTYEGTQAIYKCRGYSLGNVIV--CRKGWVALNP	58		
Db	21	ENCK--PPHFEYRVKSNTKEDLSVGETAEILICRPGVYTNKILITTECLQNGTW--STP	76		
QY	59	LRCKCRKRCRGHGDIPPEFGFTLGG-NVFEYGVKAVYTCNEGYYOLLGETINRYEC--DTD	114		
Db	77	NFCRCRRCRPPADLLNGAVHITHGDNALKFGSNTSYECNEGYDILGS--NVRFCLQDPE	135		
QY	115	--GWNDIPICVEVCKLPTAPDENKIVYSSAEMPBE-VHFOAVREYVNSGKIEGDE	171		
Db	136	NVNMDSNEVEVCIOKICIRPAVEHGDYL-----PQDYNYNGDAILTFKLSLTYLGSTT	190		
QY	172	MHCSDGDFWSKREPKVCSEICKSPVING-SPISKIITYKENERFOYKCNMGYESSERGD	230		
Db	191	LVCTSNKKWMSNFFPCLMLVCSPOIDNGYIDIGLSRRYNHQSITVCKSDYNIIVPET	250		
QY	231	AVCTESGM-RPLPSC-----EKSODNYYINGDYSPLR-----	264		
Db	251	LTCITNTWVPLPLKCVLVTNNPSTPMETPMETPTPDQKTNLSTAKTATTPNAFVTVV	310		
QY	265	-----KHRTGDEITYOCRMNGFVATRGNTAKCTSTGWI	297		
Db	311	VSPKEDVTVCKPHPEREFVKAENDKKEYSVGASVELLICRPETFTMQSTIVSECLSNGTW	370		
QY	298	PAP--RCITLKPCDYPDIKHGGLY	318		
Db	371	TAPNAKCHRRKCPPTPOELLNGEY	393		
RESULT	11				
ID	Q29530	PRELIMINARY;	PRT;	2014	AA.
AC	Q29530;				
DT	01-NOV-1996	(Tremblere1.01, Created)			
DT	01-NOV-1996	(Tremblere1.01, Last sequence update)			
DT	01-MAY-2000	(Tremblere1.13, last annotation update)			
DE	COMPLEMENT RECEPTOR 1 (FRAGMENT).				

CN	CRI.
OS	Pan troglodytes (Chimpanzee).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Pan.
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE: 94292799.
RA	Birmingham D.J., Shen X.P., Hourcade D., Nickells M.W., Atkinson J.P.;
RT	"Primary sequence of an alternatively spliced form of CRI. Candidate
RT	for the 75,000 Mr) complement receptor expressed on chimpanzee
RT	erythrocytes.";
RL	J. Immunol. 153:691-700(1994).
DR	EMBL: L24920; AAA51438.1; "-.
DR	HSSP; P08603; IHFI.
DR	INTERPRO: IPR000436; "-.
DR	INTERPRO: IPR000834; "-.
DR	INTERPRO: IPR001424; "-.
DR	PFAM: PF000084; sushi; 30.
DR	PROSITE: PS00087; SOD_CU_ZN_1; UNKNOWN_1.
DR	PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_2.
FT	NON_TER
FT	NON_TER
SO	SEQUENCE
	2014 AA; 221280 MW; 6D6C3A74D81FDB9 CRC64;

Query Match	18.6%	Score 348	DB 6	Length 2014
Best Local Similarity	28.4%	Pred. No. 1.2e-22		
Matches	96	Conservative	57	Mismatches 135; Indels 50; Gaps 23

  

Query	20	SDQTYPECTGQAIYKCRPGYRSLGNVIMV-CRKG-BWVALNPRLKCRQKRCGHPGDTPEGT	77
Db	1387	NDFEPPVSTLNTSECRPGY--FGKMFSTISCLENLVMSVED--NCRKSGCPPPE-PFNG	1441
QY	78	FTLGGANFEEYGVAKVYVYCNBSYOLGLGELNTECDTG---WTNDIPICEVYKCLPYTA	133
Db	1442	MVHINTD--QEGSTVYVYSCNBSGFRITGSDS--TTCLVSGNNVTWDRKAPICELIISCEPPT	1499
QY	134	PENCKIVYSAEPDREYHFGQAVRFVCSNG-----YKIEGDEEMHCS--DD--GFWSK	183
Db	1500	ISNGDFYNNR---ASFINGIVYVYQCCHTGPBGDLFLFELVGRSLYCTSKDQYGVWSSP	1556
QY	184	KPKCVAYS-CKSPDYINGSPT-SOKIYKENERFOYKCNMGYSESGDAVCTESG-WRP	240
Db	1557	PPRCISTKCTAPAEYEMAIRVPGNNSFFSLTEIVRFRCQPGVWMSHTVQCOTNGRMCP	1616
QY	241	-LPSCSEKSCNPYPNNDYSPLRKHK---TGEIITYQCNGNYTPATRGATYACTSG	295
Db	1617	KLPCHSRVQCPPELLHSEHP---SHQDNFSPGGEVYFSCPG-YDLRGAASLHCTPGG	1672
QY	296	-WIP-APRCTLKPCD--YPDIKHGGLVHEHMRPFPV	329
Db	1673	DWSPAPRCTYKSCDDFLGQLPHG-----RVLFPL	1702

  

RESULT	12
ID	Q29531
AC	Q29531
DT	01-NOV-1996 (Tremblrel. 01, Created)
DT	01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT	01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE	COMPLEMENT RECEPTOR 1 PRECURSOR (FRAGMENT).
GN	CR1.
OS	Pan troglodytes (Chimpanzee).
OC	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.
RN	[1]
RP	SEQUENCE FROM N.A.
RA	MEDLINE; 94292799.
RT	Birmingham D.J., Shen X.P., Hourcade D., Nickells M.W., Atkinson J.P.;
RT	"Primary sequence of an alternatively spliced form of CR1. Candidate
RT	for the 75,000 M(r) complement receptor expressed on chimpanzee
RT	erythrocytes."

Query Match  
 Best Local Similarity 29.0%; Pred. No. 9.8e-23;  
 Matches 99; Conservative 55; Mismatches 128; Indels 59; Gaps 23;

Query 3 CNE---LPPRRNTEILTGSMSDQTPPECTQAIYKCRPGYISLGNVIMVCKRGEVVALNPL 59  
 Db 18 CNAPEWLPFARPTNLT---DEFEFPITGYLNTRECRPGYGRPSIICLKNSVWTGAKD- 72  
 QY 60 RKOGRPCGHPDPPTFTLTGTVNFEYGVKAYTCNCGYQLGELINRECD---PTDG 115  
 Db 73 -RCRRKSCRNPDPVNGMWHY--KDIOFGSQIYCTKGRILGS--SSATCIISGDIYI 128  
 QY 116 WTNDIPICEVVKC-LPTVAPENGKIVSSAMEDREHFGAVRVCNSG-----YKIEG 168  
 Db 129 WDNETPICDRIPGLPPTI--TNGDFISTNRE--NFHYGSVVTYRCNCGSGRKYFELVG 184  
 QY 169 DEEMHC--SD--GFMSKERKPC-VEISCKSPDVINSPTIS-QKIIYKENERFOYKCMG 222  
 Db 185 EPSLYCTSNDDQVIGWSPAPQCIIPKCTPPNVENGILVSDNSLPSLNVVEFRCPG 244  
 QY 223 YEYSERDAVCTE-SGNRP-LPSCEEKSCDNPYIPN-----GDYSPLRIRKHRTG 269  
 Db 245 FVMKGPVRVCKQALNKKEPELPSG-SRVCORP--PDVILHARTGRDKNFSP-----G 294  
 QY 270 DEITYGCRNGFYPATRGNTAKTSTG-WIPA-PRCTIKPCD 308  
 Db 295 QEVFYSCPEG-YDLRGAASLRCTPGDMSPTPRCEVWSCD 334

RESULT 13  
 Q28797 ID Q28797 PRELIMINARY; PRT; 522 AA.  
 AC Q28797;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE UNKNOWN PROTEIN (FRAGMENT).  
 OS Pan troglodytes (Chimpanzee).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 NC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 94292799.  
 RA Birmingham D.J., Shen X.P., Hourcade D., Nickells M.W., Atkinson J.P.;  
 RT "Primary sequence of an alternatively spliced form of CrI. Candidate  
 RT for the 75,000 M(r) complement receptor expressed on chimpanzee  
 RT erythrocytes.";  
 RL J. Immunol. 153:691-700(1994).  
 DR EMBL; L24922; AAA50460.1; -.  
 DR HSSP; P08603; 1HFI.  
 DR INTERPRO; IPR000436; -.  
 DR INTERPRO; IPR000834; -.  
 DR PFAM; PF00084; sush1; 8.  
 DR PROSITE; PS00133; CARBOXYPEPT\_ZN\_2; UNKNOWN\_1.  
 FT NON\_TER 1  
 SQ SEQUENCE 522 AA; 57304 MW; DB8456EFA62C00C9 CRC64;

Query Match  
 Best Local Similarity 30.1%; Pred. No. 1.5e-22;  
 Matches 95; Conservative 52; Mismatches 119; Indels 50; Gaps 21;

Query 24 YPECTQAIYKCRPGYISLGNVIMVCKRGEVVALNPLRCKRCPGHPDPPTFTLTG 83  
 Db 5 FPIITYLNTRECRPGYGRPSIICLKNSVWTSAD--KCRKRSRNPDPVNGMWHY-- 60  
 QY 84 NFEYGVKAYVTNCGYQLGELINRECDTDG---WTNDIPICEVVKC-LPTVAPENGK 138  
 Db 61 KDIOFGSQIKYCTKGRYILGS--SSATCIISGNVIMVIMNKPVCDDRIICGLPPTI-ANGD 118  
 QY 139 IVSSAMEPDRFY-HFGAVRVCNSG-----YKIEGDEEMHC--SD--GFMSKERKPC 187  
 Db 119 FTSIS---REYFYASVVTYHCHNCGSGKRVFELVGEPSITCYSDQVIGWSPAPQC 174  
 QY 188 -VEISCKSPDVINSPTIS-QKIIYKENERFOYKCMGEYSERDAVCTE-SGNRP-LPS 243  
 Db 175 IIPKCTPPNVENGILVSDNSLPSLNVVEFRCPGVFMKGPVRVCKQALNKKEPELPS 234  
 QY 244 CEKSCDNPYIPND-----YSPLRIKHRTGDIYQCRNGFYPATRGNT-AKCTST 294  
 Db 235 CSRVCPDPVILHGERTRGRDKNFSP-----GDEVYSCPEG--DLRGSTYLHCTPQ 285  
 QY 295 G-WIP-APRCTIKPCD 308  
 Db 286 GDWSPEAPRCEVWSCD 301

RESULT 14  
 Q22328 ID Q22328 PRELIMINARY; PRT; 560 AA.  
 AC Q22328;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE COSMID T07H6.  
 CN T07H6.5.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 NC Rhabditidae; Peleoderinae; Caenorhabditis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE; 94150718.  
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,  
 RA Craxton A., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latelle P.,  
 RA Lightning J., Lloyd C., Mcmurtry A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,  
 RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,  
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans.";  
 RL Nature 368:32-38(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Geisel C.;  
 RA Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Waterston R.;  
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U53344; AAA96225.1; -.  
 DR HSSP; P10998; 1VVC.  
 DR INTERPRO; IPR000436; -.  
 DR PFAM; PF00084; sush1; 7.  
 SQ SEQUENCE 560 AA; 61619 MW; FB8923BAC1B320C9 CRC64;







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2000, 16:43:48 ; Search time 49.92 Seconds  
(without alignments)  
293.168 Million cell updates/sec

Title: US-09-316-163-14

Perfect score: 2454  
Sequence: 1 EDCKGPPRENSSEILSGSMS.....DTYCTENGMSPPKCVRIK 428

Scoring table: BLOSUM62  
Gapop 10.0 , Gapect 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

A\_Geneseq\_36:\*

- 1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT:\*
- 2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT:\*
- 3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT:\*
- 4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT:\*
- 5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT:\*
- 6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT:\*
- 7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT:\*
- 8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT:\*
- 9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT:\*
- 10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT:\*
- 11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT:\*
- 12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT:\*
- 13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT:\*
- 14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT:\*
- 15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT:\*
- 16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT:\*
- 17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT:\*
- 18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT:\*
- 19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT:\*
- 20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT:\*
- 21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	990	40.3	240	18	W39154	Human partial Comp
2	918	37.4	216	18	W39155	C10ne PRB9FN410 C
3	544.5	22.2	578	20	Y09065	Human complement f
4	481.5	19.6	581	12	R13490	Human C4 binding p
5	438	17.8	543	13	R28547	CRI-4 (52S, 53S, 5
6	437	17.8	453	20	Y55757	Human CRI protein
7	437	17.8	778	19	W73147	Amino acid sequenc
8	437	17.8	1930	19	W45899	Human complement r
9	437	17.8	2039	20	Y55751	Human C3D/C4b rece
10	437	17.8	2317	20	P92219	CRI protein. Homo
11	434	17.7	496	20	Y55752	Human CRI protein
12	434	17.7	543	13	R28557	CRI-4 (99H, 103E)

13	434	17.7	1537	12	R11982	Partial human comp
14	434	17.7	2039	14	R36743	CRI. Homo sapiens
15	432	17.6	543	13	R28570	CRI-4 (266-274 KILK
16	432	17.6	2039	12	R11810	Human complement f
17	431	17.6	450	20	Y55754	Human CRI protein
18	430	17.5	543	13	R28560	CRI-4 (114S) analo
19	429	17.5	543	13	R28548	CRI-4 (57V, 59K) a
20	428	17.4	543	13	R28550	CRI-4 (64K) analog
21	428	17.4	543	13	R28553	CRI-4 (85R, 87N) a
22	428	17.4	543	13	R28558	CRI-4 (109N, 110A,
23	428	17.4	543	13	R28565	CRI-4 (121Q) analo
24	428	17.4	543	13	R28571	CRI-4 (364-367 NAA
25	427	17.4	543	13	R28545	CRI-4 (37Y) analog
26	427	17.4	543	13	R28552	CRI-4 (78T, 79D) a
27	426	17.4	543	13	R28555	CRI-4 (92T) analog
28	426	17.4	543	13	R28563	CRI-4 (117P) analo
29	426	17.4	543	13	R28567	CRI-4 (318-321 RNP
30	425	17.3	543	13	R28549	CRI-4 (64K, 65T) a
31	425	17.3	543	13	R28551	CRI-4 (65T) analog
32	425	17.3	543	13	R28552	CRI-4 (116K) analo
33	425	17.3	543	13	R28568	CRI-4 (347T, 349Y)
34	424	17.3	543	13	R28569	CRI-4 (369-376 STK
35	424	17.3	543	13	R28566	CRI-4 (94H) analog
36	423	17.2	543	13	R28564	CRI-4 (116K, 117P)
37	422	17.2	543	13	R28561	CRI-4 (115T) analo
38	421	17.2	543	13	R28554	CRI-4 (92T, 94H) a
39	420	17.1	543	13	R28546	CRI-4 (35E) analog
40	420	17.1	543	13	R28546	CRI-4 (44T, 47P, 4
41	420	17.1	543	13	R28566	CRI-4 (318R, 319N)
42	419	17.1	543	13	R28543	CRI-4 (35E, 37Y) a
43	419	17.1	543	13	R28559	CRI-4 (114-115TKP
44	412	16.8	44	20	Y55756	Human CRI protein
45	409	16.7	450	20	Y55753	Human CRI protein

## ALIGNMENTS

RESULT 1	
W39154	W39154 standard; Protein; 240 AA.
AC	W39154;
XX	
DT	27-APR-1998 (first entry)
DE	Human partial Complement factor H protein fragment 1.
XX	
KW	Complement factor H; tumour associated antigen; renal cancer; urogenital cancer; medicament; modulator.
XX	
OS	Homo sapiens.
PN	WO9738136-A1.
XX	
PD	16-OCT-1997.
XX	
PF	09-APR-1997; 97WO-US05710.
XX	
PR	06-MAR-1997; 97US-0812481.
PR	09-APR-1996; 96US-0015083.
PR	09-APR-1996; 96US-0630048.
PR	06-MAR-1997; 97US-0038614.
PA	(BARD-) BARD DIAGNOSTIC SCI INC.
PI	Enfield DL, Hass GM, Kinders RJ;
XX	
DR	WPI: 1997-512742/47.
DR	N-PSDB: V02790.
XX	
PT	Treating or screening for cancer, e.g. renal or urogenital cancer - by modulating or detecting tumour associated human complement factor

PT H related antigen, or nucleic acid encoding it  
 PS Example 6B; Fig 6B; 104pp; English.

CC This partial protein sequence represents a region of the human  
 CC tumour-associated complement factor H (CFH). This sequence is used  
 CC in the identification of complement factor H related proteins and  
 CC antigens isolated from clone pRBB9FH410 (see W39155). The detection of  
 CC such proteins and a CFH antigen can be used in screening or for the  
 CC treatment of renal or urogenital cancer, e.g. bladder, cervical or  
 CC prostate cancer. Agents that may modulate this antigen could be used in  
 CC the manufacture of a medicament for the treatment of a tumour cell.

SO Sequence 240 AA;

Query Match 40.3%; Score 990; DB 18; Length 240;  
 Best Local Similarity 71.7%; Pred. No. 3.2e-68;  
 Matches 172; Conservative 21; Mismatches 47; Indels 0; Gaps 0;

QY 78 FRNAVSEFEFGAKVYTCDEGQLGELIDYRECDADGWTNDIPICEVVKCLPTELENG 137  
 DB 1 fllggnvteygvkavycnecyqllelnyrecdlgwntndipicevkcclptapeng 60  
 QY 138 RIVSGAEPDQEVYFGOVVRFECNSGFKIEGOKEMHCSEGLMSNEKPCQVEISCLPPRY 197  
 DB 61 kvssamepdeyhfqgavrfvcnsygllegdeemcsddgfwkpkcveiscspdv 120  
 QY 198 ENGDDGYLKPVEKNERFOYKCKQGFVYKERGDVCTGSGWNPQSCSEMTCLTPYIPNG 257  
 DB 121 lngspisqkilykenerfykcmgyeysergdavctesgwrlpseeckscdnpylpng 180  
 QY 258 IYTPRIKIRIDETRECKNGFYPRTRSPVSKCTTGTGTPARCSLKCDDPFQFNGRL 317  
 DB 181 dysplriktrtcdeltgcrngfpatrgntactstgwlipaprclickpodypdikhg91 240

RESULT 2

ID W39155 standard; Protein; 216 AA.

AC W39155;

DT 27-APR-1998 (first entry)

DE Clone pRBB9FH410 CFH related protein fragment.

KW Complement factor H; tumour associated antigen; renal cancer;

KW urogenital cancer; medicament; modulator.

OS Synthetic.

PN WO9738136-A1.

PD 16-OCT-1997.

PF 09-APR-1997; 97WO-US05710.

PR 06-MAR-1997; 97US-0812481.

PR 09-APR-1996; 96US-0015083.

PR 09-APR-1996; 96US-0630048.

PR 06-MAR-1997; 97US-0038614.

PA (BARD-) BARD DIAGNOSTIC SCI INC.

PI Enfield DL, Haas GM, Kinders RJ;

DR WPI: 1997-512742/47.

DR N-PSDB; V02791.

PT Treating or screening for cancer, e.g. renal or urogenital cancer -  
 PT by modulating or detecting tumour associated human complement Factor  
 PT H related antigen, or nucleic acid encoding it

XX  
 PS Example 6B; Fig 6B; 104pp; English.

CC This partial protein is found in clone pRBB9FH410 and represents a  
 CC complement factor H related protein with homology to a region of the  
 CC human tumour-associated complement factor H (CFH). The detection of this  
 CC protein and a CFH antigen can be used in screening or for the treatment  
 CC of renal or urogenital cancer, e.g. bladder, cervical or prostate cancer.  
 CC Agents that may modulate this antigen could be used in the manufacture of  
 CC a medicament for the treatment of a tumour cell.

SO Sequence 216 AA;

Query Match 37.4%; Score 918; DB 18; Length 216;  
 Best Local Similarity 73.1%; Pred. No. 8.8e-63;  
 Matches 158; Conservative 19; Mismatches 39; Indels 0; Gaps 0;

QY 95 TCDEGYQLGELIDYRECDADGWTNDIPICEVVKCLPTELENGRIVSGAEPDQEVYFG 154  
 DB 1 tconeyqlllgelnyrecdlgwntndipicevkcclptapengkivssamepdeyhfqg 60  
 QY 155 VVRFECNSGFKIEGOKEMHCSEGLMSNEKPCQVEISCLPPRYENGDTYKRYKENER 214  
 DB 61 avrfvcnsygllegdeemcsddgfwkpkcveiscspdvlnsgspisqkilykener 120  
 QY 215 FOYKCKQGFVYKERGDVCTGSGWNPQSCSEMTCLTPYIPNGIYTPRIKIRIDETIRY 274  
 DB 121 fgykcmgyeysergdavctesgwrlpseeckscdnpylpngdysplriktrtcdeltg 180  
 QY 275 ECKNGFYPRTRSPVSKCTTGTGTPARCSLKCDDPF 310  
 DB 181 qcrngfpatrgntactstgwlipaprclickpodyp 216

RESULT 3

ID Y09065 standard; Protein; 578 AA.

AC Y09065;

DT 06-JUL-1999 (first entry)

DE Human complement factor H homolog protein.

KW Human complement factor H; immunological mechanism; complement reaction;

KW gene therapy; immune stimulation; haematopoiesis regulation; chemotactic;

KW tissue growth activity; anti-inflammatory; tumour inhibition;

KW secretory signal.

OS Homo sapiens.

PN WO9918200-A1.

PD 15-APR-1999.

PF 02-OCT-1998; 98WO-JP04448.

PR 06-OCT-1997; 97JP-0272837.

PA (PROT-) PROTEGENE INC.

PA (SAGA) SAGAMI CHEM RES CENT.

PI Kato S, Sekine S;

DR WPI: 1999-264019/22.

DR N-PSDB; X34737.

PT Human proteins with secretory signal sequences and nucleotide  
 PT sequences, useful in control of proliferation and differentiation of  
 PT cells  
 PS Claims 1; Page 55-58; 71pp; English.

XX This DNA encodes a protein having homology to human complement factor H,  
 CC which plays a role in the immunological mechanism involving the  
 CC complement reaction. The protein can also be used as an antigen for  
 CC preparing antibodies against the protein. The cDNA can be used as a probe  
 CC for gene diagnosis and the gene for gene therapy, as well as for large-  
 CC scale expression of the protein. The protein may also have immune  
 CC stimulating or suppressing activity, haematopoiesis regulating activity,  
 CC tissue growth activity, activin/inhibin activity, anti-inflammatory  
 CC activity, tumour inhibition activity, chemotactic/chemokinetic activity,  
 CC receptor/ligand activity, etc. The protein is identified by the presence  
 CC of a hydrophobic N-terminal secretory signal region, knowledge of the  
 CC protein function is not required, as in e.g. methods of expression  
 CC cloning.

SO Sequence 578 AA;

Query Match 22.2%; Score 544.5; DB 20; Length 578;  
 Best Local Similarity 32.3%; Pred. No. 9,1e-34;  
 Matches 136; Conservative 55; Mismatches 147; Indels 83; Gaps 24;

48 CRKNEWVSPNSRRCRRPCGHP-----GDPFGSFR-----LAVGSEFEFGAKVVTCD 98  
 15 canqev-----kpcdfpeiqhglykylslrlyfpaaqgys-----yycdq 58  
 99 GYOL-LIGEI-DYRECDADGWTNDIPICEVVKLPV-----TELENGRIVSGAAEPDOEY 151  
 59 nfvfpgsgsywdylhctqdgwspv-----clrtcskdsdvelngfl-----sessilyl 108  
 152 FGOVVRRECSNGFKI-----EOKEMHCSENGLSMNEKPCQVEISLPPRVEN-----GDGIV 204  
 109 ineetycnckpyatadgnssgsltcldqng-wslq-piclkf-cdmvfensraksngmw 165  
 205 LKPYKKEKNEROYKCKGKF--VYKERGDA--VCTSGWNNPQPSC--DEMTCLTPYIPNG- 257  
 166 -----fkhdtdldecydgysyngntdsivcgedgshlptcynssescgpppisingd 221  
 258 -----IYTPHRIKRIHIDIRYECCKNGEYFPAIRSPVSKCTITGWIAPRC-SLKPCDF 309  
 222 ttfspqkylpw-----srveyqcs-yelqgskylvtsnsgdwsepprcismkpccef 273  
 310 POFHGRILYRESRPYFPVPTGKEYSYCDNGTTTSQSQWDLRCTVNGMEFEVPCLR 369  
 274 peidghlyhyentrryfpvatqgysyycdqntfvtpsgwydylhctqdgwlpvlpcl 333  
 370 QCIFH--VVEGESSYVQRRYIEGOSAKVOCBSGYSLPNGD--TYVCTENGSMSPPKCV 425  
 334 tcsksdleiengfissessylilnkelqykckpyatadgnssgsltcldqngwsaqpdlcl 393  
 426 R 426  
 394 k 394  
 Db  
 RESULT 4  
 R13490  
 ID R13490 standard; Protein; 581 AA.  
 XX  
 AC R13490;  
 XX  
 DT 30-OCT-1991 (first entry)  
 XX  
 DE Human C4 binding protein.  
 XX  
 KM C4bp: monomer; complement protein; pJOD.C4bp.3; SCR;  
 XX short consensus repeat.  
 OS Homo sapiens.  
 XX  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..32  
 FT label= signal\_peptide

FT Protein 33..581  
 FT /label= C4bp  
 FT Region 33..93  
 FT /label= SCR8  
 FT Region 94..155  
 FT /label= SCR7  
 FT Region 156..219  
 FT /label= SCR6  
 FT Region 220..279  
 FT /label= SCR5  
 FT Region 280..345  
 FT /label= SCR4  
 FT Region 346..406  
 FT /label= SCR3  
 FT Region 407..464  
 FT /label= SCR2  
 FT Region 465..523  
 FT /label= SCR1  
 FT Domain 524..581  
 FT /label= C4bp\_core  
 FT /note= "responsible for multimer assembly"  
 FT Disulfide-bond 34..80  
 FT /note= "intradomain"  
 FT Disulfide-bond 65..92  
 FT /note= "intradomain"  
 FT  
 FT WO9111461-A.  
 FT  
 FT 08-AUG-1991.  
 FT  
 FT 28-JAN-1991; 91WO-US00567.  
 FT  
 FT 26-JAN-1990; 90US-0470888.  
 FT  
 FT (BIOG-) BIOGEN INC.  
 FT  
 FT Pasak MP, Winkler G, Liu TR;  
 FT  
 FT WPI; 1991-252613/34.  
 FT  
 FT N-PSDB; Q13242.  
 FT  
 FT New C4 binding protein fusion proteins and DNA encoding them -  
 FT comprise assemblies of C4bp monomers linked to functional moiety,  
 FT e.g. AZT, useful as delivery vehicles in diagnosis and therapy  
 FT  
 FT Example 1; Fig 1; 105pp; English.  
 FT  
 FT This sequence was deduced from human hepatocyte (Hep G2) cDNA  
 FT obtained following PCR amplification. The protein is a monomer  
 FT containing 8 SCRs. Each SCR forms a looped domain due to the  
 FT presence of two intradomain disulphide bonds (only the disulphide  
 FT bonds of SCR8 are labelled in the Features Table). Within each SCR,  
 FT the first cysteine residue bonds with the third and the second  
 FT cysteine residue bonds with the fourth. This secondary structure is  
 FT responsible for the conformational flexibility of the C4bp monomer.  
 FT The invention covers fusion proteins in which the monomer sequence,  
 FT or a specified subfragment of it, e.g. having 5, 4, 3 or 1 SCR(s)  
 FT is fused to the C-terminal of a protein such as a viral receptor,  
 FT cell ligand, a bacterial, viral or parasitic immunogen, enzyme,  
 FT cytokine, toxin, etc. See also Q13243-51.  
 FT  
 FT SO Sequence 581 AA;

Query Match 19.6%; Score 481.5; DB 12; Length 581;  
 Best Local Similarity 26.5%; Pred. No. 5,8e-29;  
 Matches 134; Conservative 78; Mismatches 195; Indels 99; Gaps 25;  
 2 DCKGPPPRENSSEISGMSQISSEGTQATYKCRPGY-RTLGITIVKVC-KNGEWPSPNS 59  
 33 ncpplstfaampdlilteerritgttlkytlpvyrvshsqtlctnsdsgwynt-- 90  
 60 RICRRRCGHPGDTFPGSFLAVGSEFEFGAKVVTCDGCVOLLGELIDYRECDAD----G 115

```

Db 91 -fclykrcthrppglnng--qveiktdlstrsgqlfscsegfllgltstr--cewqdgvgv 146
OY 116 WTMNDIPICEVVKCLPTELENGRIVSGAEPDDEYVGVVREECNSGFKIEGOKEMHCS 175
Db 147 wshp1bpceivkckppddlngrf----hsgeenfyaygfvtyscdprflshasisc 202
OY 176 -EN---GIMSNEKPOCVELISCLPRVENGGIT-LKPVYKENERFOYKCKOGFVYKERGD 230
Db 203 venetlgvwrpspplceklctcrkpdvshgemwsgfiplynkdtllyfkcgkgtv1--rgs 260
OY 231 AV--C-TGSQWNPQ-PSCEMTCLT-PIYPNGIYT--PHRIK--HRIDEIRYECKNGF 280
Db 261 svlhcdadckmpspaccenscinlpdiphaswetyprptkedyyvvgvltvlyrcmipy 320
OY 281 YRATRSVPVKCTITG--WIPAPRCSLKPCDFQPKHGRLYVESRRP----- 325
Db 321 kpltdptlvicqknltwrlpygcealcepepklmgeltqhrkstrpanhcvfygdeis 380
OY 326 -----YRPVPIG-----KEYSYCDN 341
Db 381 fchetsrfsalcqdgqvtwspitpsscglcnlfpkiahghyqssysffkeelylecdk 440
OY 342 GFTTSPQSYWDYLRCTVNGMEPEVP-CLRQCIFHYVEYESSYWRRTYEGOSAKYQCHS 400
Db 441 gyllyvga---klscsyshwsapapqckalcrkpelvnglrlsvdkqgyevpenvtlqcds 497
OY 401 GYSLRNGDITYCTEN-GMSPP-PKC 424
Db 498 gylgvv-gpqsitcsqurtwlypevpkc 522

```

```

RESULT 5
R28547
ID R28547 standard; peptide: 543 AA.
XX R28547;
DT 19-MAR-1993 (first entry)
XX CRI-4 (52S, 53S, 54P) analogue.
DE short consensus repeat; regulator of complement activation;
KW C3b binding; C4b binding; human complement type 1 receptor.
XX Homo sapiens.
XX OS
XX FH
XX Key Location/Qualifiers
XX Region 1..60
XX Region /label= SCR-1
XX Region 61..122
XX Region /label= SCR-2
XX Region 451..510
XX Region /label= SCR-8
XX Region 511..543
XX Region /label= SCR-9
XX Region /note= "TRUNCATED"
XX Misc-difference 52
XX FT /note= "Thr substituted by Ser (SCR-8)"
XX FT Misc-difference 53
XX FT /note= "Gly substituted by Ser (SCR-8)"
XX FT Misc-difference 54
XX FT /note= "Ala substituted by Pro (SCR-8)"
XX
XX EP512733-A.
XX
XX 11-NOV-1992.
XX
XX 28-APR-1992; 92EP-0303826.
XX
XX 03-MAY-1991; 91US-0695514.
XX
XX (UNIW ) UNIV WASHINGTON.

```

```

XX
PI Atkinson JP, Hourcade D, Krych M;
XX
XX WPI; 1992-375009/46.
XX
PT Complement activity regulator protein analogues - useful for
PT treating auto-immune diseases, to suppress transplant rejection,
PT for diagnosis etc.
XX
PS Claim 11; Fig 2 and R11810; 23pp; English.
XX
CC The CDNA clone designated CRI-4 was described in J. Exp. Med. (1988)
CC 168:1255-1270. It encodes the first 8 and a half amino terminal
CC SCRs of CRI. The invention concerns analogues of "regulator of
CC complement activation" proteins or truncated, hybrid or recombinant
CC forms of them. CRI-4 is a preferred truncated form and a number of
CC specified substitution variants of it are claimed in which certain
CC positions in SCR-1 which have been identified as important for the
CC degree of C3b- and C4b-binding are substituted by amino acids from
CC the corresponding positions in SCR-8. The specification does not
CC contain the CRI-4 sequence; the sequence given here was constructed
CC from the full-length CRI amino acid sequence having GEMESBQ
CC accession number R11810 and descriptions in the disclosure.
XX
SQ Sequence 543 AA;

```

```

Query Match 17.8%; Score 438; DB 13; Length 543;
Best Local Similarity 27.9%; Pred. No. 1,le-25;
Matches 133; Conservative 63; Mismatches 148; Indels 132; Gaps 31;

```

```

OY 27 GTQATYKCRPGRYTGLTIVKCKNGEWPNSBRLCRKRCGHPGDTPGSRLAVSEF 86
Db 25 gtylnyecrpyysgrpfslclksnw--spskdrctrcrscnpprvngwlvikg--1 80
OY 87 EFGAKVVTCTDGGYOLGIEDYRECDAGS---WTNDIPICEVVKCLPTELENGRIYS 141
Db 81 qfsgqtkysctgyrlilgs-satclisgdtvawdneipcdrlpcglpdt-iltgdtis 138
OY 142 GAEPDQEYVFGQVNFECNSG-----FKIEGOKEMHCSEN---GIMSNEKPOC-VBI 190
Db 139 tnre--nfhygsvlyrcnpgsgarkvfelvgpeiyctsnddqyivsgapqclipn 195
OY 191 SCLPRVENG---DGIVLKPYKENERQYKCKOGFYVKKENGDAVCTG-SCWNPQ-PSC 244
Db 196 kctppvengllvsdn---tslfsinevelfrcqpylmykpryvkcgqlnkwepe1psc 252
OY 245 EEMTCLTYPIPIYNGIYTPHRIKHRIID---DEIRYECKNGFYPATRSPVSKCTITG-WIP- 298
Db 253 svr-cqpp--pdlvlnaertqrkdnlfpqgevfycsepg-ydlirgaasmrctprgdwspa 308
OY 299 ARCSLSKPCD--FPOFKHGRLYVESRRPFPV--PIGKEYSYVDCNGFTT--PSQSYWD 352
Db 309 aptcevkscddfmqllngvr-----lfpvnlqlyakvdfvcdqfqlkgsasy-- 358
OY 353 YLRCTVNG-----WEPEVPCLRQCIFHYVEYESSYWRRTYEGOSAKYQCHGYSLPN-- 406
Db 359 ---cvlagmeslmsvpyceq-----ifcspvvlmgtr 390
OY 407 -----GDDTYVCTE-----NG-W-SPPPKC 424
Db 391 htgkplevfpfgkavnyctcdphdrtgtsfdllgestlctsdpgngywsppaprc 446

```

```

RESULT 6
Y55757
ID Y55757 standard; Protein; 453 AA.
XX
XX Y55757;
XX
XX 22-FEB-2000 (first entry)
XX
XX Human CRI protein LHR-C fragment.
DE

```



[illegible]



```

Db 75 gtylnyecrpgysgrfslclksnwtgkadr--crkscsrnpdpdvngmwhikg--i 130
QY 87 EFGAKVYVTCDEGQOLGELIDYRCADG---WTNDIPICEVVC-LPYTELENGRIVS 141
Db 131 qfsgqikysctkyrligs--ssatclisgdtvldnetpdrpcglppl-ltngdfl 188
QY 142 GAAPDDEYFGOVVRECEGSG---FKIEGQKEMHCSEN---GLMSNEKPOC-VEI 190
Db 189 tne-----nfhygsvvtlycnpqsggrkvfclvegepslycsnddqvgylwspapqcllpn 245
QY 191 SCLPRVENG---DGITYLKPVYKEMERFQYKCKOGFYVKEGDAVCTG--SGWNPQ--PSC 244
Db 246 kcpvnevgilvsdn---zslfslnevefrcqpfvmkgrprvkcqalnkewelpsc 302
QY 245 EEMTCLLPYIPNGIYPRHRIKHRIID---DEIRYCKNGFYPRATRSVSKCTITG-WIP 298
Db 303 srvcqpp--pdvlnhaertgrdkdnfsgqevfyscepg-ydlrgaasmrctpgdwsa 358
QY 299 APRCSLKPCD--PPOFKHGRLYEESRRPYPV--PIGKEYSYCDNGFTT--PSQSYMD 352
Db 359 apcevkscddfmqgllntrv-----lfvnlqigakvdtvcdedqfqlkgsasay-- 408
QY 353 YLRCTVNG---WEPEVPCLRQCIFHYVEYEGESSYQWRXYIEGOSAKVQCHSGYSLPN-- 406
Db 409 ---cvlagmeslwnssvpcq-----lfcpvprvlpnqr 440
QY 407 -----GODTYCTE---NG-W-SPPK 424
Db 441 hngkplevfifgkavnytcphdpdrgtsfdlqgestlrcsdpgngvwspparc 496

RESULT 11
Y55752
ID Y55752 standard; Protein: 496 AA.
AC Y55752;
DE 22-FEB-2000 (first entry)
XX Human CRI protein LHR-A SCR fragment.
XX C3B/C4B receptor; CRI protein; cell-surface protein; erythrocyte; human;
XX complement regulatory activity; complement pathway enzyme; tissue damage;
XX reperfusion injury; Arthus reaction; myocardial infarct; inflammation;
XX heart condition; autoimmune disorder; long homologous repeat; LHR; SCR;
XX short consensus repeat.
XX Homo sapiens.
XX OS
XX US5981481-A.
XX
XX 09-NOV-1999.
XX
XX 06-JUN-1995; 95US-0470652.
XX
XX 03-APR-1989; 89US-0332865.
XX 06-DEC-1974; 74US-0350238.
XX 24-FEB-1993; 93US-0026134.
XX 01-APR-1988; 88US-0176532.
XX
XX (UYJO ) UNIV JOHNS HOPKINS.
XX (BGHM ) BRIGHAM & WOMENS HOSPITAL.
XX (AVAN-) AVANT IMMUNOTHERAPEUTICS INC.
XX
XX Conclno MF, Wong WM, Makrides SC, Klickstein LB, Fearon DT, IP SH;
XX Marsh HC, Carson GR;
XX WPI; 1999-633357/54.
XX A human C3B/C4B receptor (CRI) protein having antiinflammatory and
XX cardiant activity -
XX

```

```

PS Disclosure; Fig 10; 87pp; English.
XX
CC The invention relates to a human C3B/C4B receptor (CRI) protein. The CRI
CC protein or fragment is expressed as a cell-surface protein on the surface
CC of a non-human cell and exhibits a complement regulatory activity of full
CC -length human CRI and exhibits a complement regulatory activity of full
CC length human CRI as expressed on erythrocytes. The CRI function in vivo
CC may be mediated through the inhibition of complement pathway enzymes. The
CC soluble CRI protein exhibits a complement regulatory activity, and this
CC may be used to prevent reperfusion injury, inhibit Arthus reaction, and
CC neutrophil mediated tissue damage, and reduce myocardial infarct size,
CC and inflammation. The CRI protein and its fragments can also be used in
CC the treatment of conditions which involve unwanted complement activity,
CC e.g. shock lung, tissue damage due to burn, or ischemic heart conditions,
CC and autoimmune disorders. CRI proteins, analogues, derivatives, and anti-
CC -CRI antibodies are used in assays, and diagnostics. The present sequence
CC represents the short consensus repeat (SCR) fragments of human CRI
CC protein long homologous repeat (LHR)-A sequence.
XX
SQ Sequence 496 AA:

Query Match 17.7%; Score 434; DB 20; Length 496;
Best Local Similarity 27.8%; Pred. No. 2e-25;
Matches 133; Conservative 63; Mismatches 148; Indels 134; Gaps 32;

QY 27 GTQATYKCRPGYRTLTGTYVVC-KNGEWPNSPRICRKRPGDTPGSRFLAVGSE 85
Db 71 gtylnyecrpgysgrfslclksnwtgkadr--crkscsrnpdpdvngmwhikg-- 126
QY 86 FEEGAKVYVTCDEGQOLGELIDYRCADG---WTNDIPICEVVC-LPYTELENGRIV 140
Db 127 idfgsqikysctkyrligs--ssatclisgdtvldnetpdrpcglppl-ltngdfl 184
QY 141 SGAAPDDEYFGOVVRECEGSG---FKIEGQKEMHCSEN---GLMSNEKPOC-VEI 189
Db 185 stnre---nfhygsvvtlycnpqsggrkvfclvegepslycsnddqvgylwspapqcllp 241
QY 190 ISCLPRVENG---DGITYLKPVYKEMERFQYKCKOGFYVKEGDAVCTG--SGWNPQ--PS 243
Db 242 nkctprvnevgilvsdn---zslfslnevefrcqpfvmkgrprvkcqalnkewelp 298
QY 244 CEEMTCLLPYIPNGIYPRHRIKHRIID---DEIRYCKNGFYPRATRSVSKCTITG-WIP 298
Db 299 csrv-cqpp--pdvlnhaertgrdkdnfsgqevfyscepg-ydlrgaasmrctpgdwsa 354
QY 299 -APRCSLKPCD--PPOFKHGRLYEESRRPYPV--PIGKEYSYCDNGFTT--PSQSY 350
Db 355 aaplccevkscddfmqgllntrv-----lfvnlqigakvdtvcdedqfqlkgsasay 406
QY 351 WDYLCTVNG---WEPEVPCLRQCIFHYVEYEGESSYQWRXYIEGOSAKVQCHSGYSLPN 406
Db 407 -----cvlagmeslwnssvpcq-----lfcpvprvlpn 436
QY 407 -----GODTYCTE---NG-W-SPPK 424
Db 437 grhtgkplevfifgkavnytcphdpdrgtsfdlqgestlrcsdpgngvwspparc 494

RESULT 12
R28557
ID R28557 standard; peptide; 543 AA.
AC R28557;
XX
XX 19-MAR-1993 (first entry)
XX
XX CRI-4 (99H, 103E) analogue.
XX
XX Short consensus repeat; regulator of complement activation;
XX C3b binding; C4b binding; human complement type 1 receptor.
XX
XX Homo sapiens.
XX

```







---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2000, 16:47:40 ; Search time 47.7 Seconds

(Without alignments)  
569,412 Million cell updates/sec

Title: US-09-316-163-14

Perfect score: 2454  
Sequence: 1 ECKCKGPPRENSSEILSGSWS.....DIYYCTENGWSPPKCVRIK 428

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters: 182106

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : PIR.65:\*  
1: pirl:\*  
2: pirl:\*  
3: pirl:\*  
4: pirl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2055	83.7	1234	1	NBM5H	complement factor
2	1714	69.8	1231	1	NBH0H	complement factor
3	1709	69.6	449	1	NBH0H	complement factor
4	1242	50.6	669	2	S6551	factor H - bovine
5	813	33.1	452	2	A35068	complement factor
6	811	33.0	303	2	H35068	apolipoprotein H-r
7	569.5	23.2	1053	2	S46199	probable complement
8	566	23.1	808	2	D35069	complement factor
9	529.5	21.6	597	1	S53711	C4BP alpha chain p
10	481.5	19.6	597	1	NBH0C4	C4BP binding protei
11	474.5	19.3	610	1	I46001	C4b-binding protei
12	466.5	19.0	331	2	A45222	complement factor
13	453	18.5	676	2	A45900	complement C3b rec
14	449	18.3	558	2	S57953	C4BP protein alpha
15	447	18.2	2014	2	I36936	complement recepto
16	437	17.8	2489	2	I73012	complement C3b/C4b
17	409	16.7	497	2	JC2054	complement regulat
18	400.5	16.3	482	2	A34924	complement C3b/C4b
19	396	16.1	579	2	A56740	sperm-egg recogniti
20	384.5	15.7	469	1	NBM5C4	C4b-binding protei
21	375	15.3	560	2	T16833	hypothetical protei
22	373	15.2	1025	1	A43526	complement C3d/Bps
23	363	14.8	1091	1	PL0009	complement C3d/Bps
24	336	13.7	661	1	KFH013	coagulation factor
25	332	13.5	668	2	A46013	coagulation factor
26	321	13.1	433	2	A30550	complement C3b/C4b
27	319	13.0	443	2	G35070	apolipoprotein H-r
28	318	13.0	440	2	A43519	complement recepto
29	315	12.8	330	2	I56100	complement factor

30	314.5	12.8	263	1	WMV2SP	apolipoprotein H h
31	310.5	12.7	768	2	A42755	P-selectin precurs
32	310	12.6	270	2	I37278	complement factor
33	309.5	12.6	830	2	A30359	P-selectin precurs
34	298.5	12.2	377	2	I54479	membrane cofactor
35	298.5	12.2	384	2	S01896	membrane cofactor
36	297.5	12.1	263	1	C36838	complement control
37	297.5	12.1	263	2	T28450	hypothetical prote
38	296.5	12.1	263	2	B72152	B18L protein - var
39	296	12.1	768	2	I53821	P-selectin - rat
40	294.5	12.0	362	2	JC5194	membrane cofactor
41	294.5	12.0	369	2	JC5138	membrane cofactor
42	293.5	12.0	369	2	G02913	sperm CD46 - human
43	293.5	12.0	369	2	I57998	membrane cofactor
44	290.5	11.8	360	2	T42921	complement control
45	290	11.8	610	2	A35046	E-selectin precurs

## ALIGNMENTS

RESULT 1  
NBM5H  
complement factor H precursor - mouse  
N:Alternate names: protein beta-1-H  
C:Species: Mus musculus (house mouse)  
C:Date: 30-Sep-1987 #sequence\_revision 30-Sep-1987 #text\_change 22-Jun-1999  
C:Accession: A26154; I49711; I49728  
R:Kristensen, T.; Tack, B.F.  
Proc. Natl. Acad. Sci. U.S.A. 83, 3963-3967, 1986  
A:Title: Murine protein H is comprised of 20 repeating units, 61 amino acids in length  
A:Reference number: A26154; MUID:8623353  
A:Accession: A26154  
A:Molecule type: mRNA  
A:Residues: 1-1234 <KRI>  
A:Cross-references: GB:M12660; NID:G193724; PID:AAA37759.1; PID:G387181  
J: Natsume-Sakai, S.; Nonaka, M.; Nonaka, M.; Harada, Y.  
J. Immunol. 144, 358-362, 1990  
A:Title: Demonstration of an unusual allelic variation of mouse factor H by the compl  
A:Reference number: I49711; MUID:90111033  
A:Accession: I49711  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-18 <RES>  
A:Cross-references: GB:M31979; NID:G193726; PID:AAA37762.1; PID:G193729  
R: Munoz-Canooves, P.; Tack, B.F.; Viik, D.P.  
Biochemistry 28, 9891-9897, 1989  
A:Title: Analysis of complement factor H mRNA expression: Dexamethasone and IFN-gamma  
A:Reference number: I49728; MUID:90148935  
A:Accession: I49728  
A:Molecule type: mRNA  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Residues: 1-19 <RE2>  
A:Molecule type: mRNA  
A:Cross-references: GB:J02891; NID:G193805; PID:AAA37795.1; PID:G553926  
C:Comment: Two codominant alleles of factor H are present in mice.  
C:Comment: Factor H functions as a cofactor in the inactivation of C3b by serine prot  
(C5 convertase) in the alternative complement pathway.  
C:Genetics:  
A:Map position: 1  
C:Superfamily: complement factor H; complement factor H repeat homology  
C:Keywords: complement alternate pathway; duplication; glycoprotein; plasma  
F:1-18/Domain: signal sequence #status predicted <SIG>  
F:19-1234/Product: complement factor H #status predicted <MP1>  
F:21-80/Domain: complement factor H repeat homology <FH01>  
F:85-144/Domain: complement factor H repeat homology <FH02>  
F:146-205/Domain: complement factor H repeat homology <FH03>  
F:210-262/Domain: complement factor H repeat homology <FH04>  
F:246-248/Region: cell attachment (R-G-D) motif  
F:325-385/Domain: complement factor H repeat homology <FH05>  
F:389-442/Domain: complement factor H repeat homology <FH06>  
F:448-505/Domain: complement factor H repeat homology <FH08>  
F:509-564/Domain: complement factor H repeat homology <FH09>

F:569-622/Domain: complement factor H repeat homology <FH10>  
 F:629-683/Domain: complement factor H repeat homology <FH11>  
 F:590-743/Domain: complement factor H repeat homology <FH12>  
 F:752-802/Domain: complement factor H repeat homology <FH13>  
 F:808-861/Domain: complement factor H repeat homology <FH14>  
 F:867-931/Domain: complement factor H repeat homology <FH15>  
 F:936-989/Domain: complement factor H repeat homology <FH16>  
 F:994-1048/Domain: complement factor H repeat homology <FH17>  
 F:1053-1107/Domain: complement factor H repeat homology <FH18>  
 F:1114-1168/Domain: complement factor H repeat homology <FH19>  
 F:1172-1233/Domain: complement factor H repeat homology <FH20>  
 F:21-66,52-88,85-129,114-141,146-192,178-205,210-251,237-262,267-309,294-320,325-374,357-402,808-850,836-861,867-920,906-931,936-978,964-989,994-1037,1023-1048,1053-1096,1082-1107,1114-1168,1172-1233,1237-1262,1267-1309,1314-1341,1346-1392,1400-1448,1453-1501,1506-1554,1559-1607,1612-1660,1665-1713,1718-1766,1771-1819,1824-1872,1877-1925,1930-1978,1983-2031,2036-2084,2089-2137,2142-2190,2195-2243,2248-2296,2301-2349,2354-2402,2407-2455,2460-2508,2513-2561,2566-2614,2619-2667,2672-2720,2725-2773,2778-2826,2831-2879,2884-2932,2937-2985,2990-3038,3043-3091,3096-3144,3149-3197,3202-3250,3255-3303,3308-3356,3361-3409,3414-3462,3467-3515,3520-3568,3573-3621,3626-3674,3679-3727,3732-3780,3785-3833,3838-3886,3891-3939,3944-3992,3997-4045,4050-4098,4103-4151,4156-4204,4209-4257,4262-4310,4315-4363,4368-4416,4421-4469,4474-4522,4527-4575,4580-4628,4633-4681,4686-4734,4739-4787,4792-4840,4845-4893,4898-4946,4951-4999,5004-5052,5057-5105,5110-5158,5163-5211,5216-5264,5269-5317,5322-5370,5375-5423,5428-5476,5481-5529,5534-5582,5587-5635,5640-5688,5693-5741,5746-5794,5799-5847,5852-5900,5905-5953,5958-6006,6011-6059,6064-6112,6117-6165,6170-6218,6223-6271,6276-6324,6329-6377,6382-6430,6435-6483,6488-6536,6541-6589,6594-6642,6647-6695,6700-6748,6753-6801,6806-6854,6859-6907,6912-6960,6965-7013,7018-7066,7071-7119,7124-7172,7177-7225,7230-7278,7283-7331,7336-7384,7389-7437,7442-7490,7495-7543,7548-7596,7601-7649,7654-7702,7707-7755,7760-7808,7813-7861,7866-7914,7919-7967,7972-8020,8025-8073,8078-8126,8131-8179,8184-8232,8237-8285,8290-8338,8343-8391,8396-8444,8449-8497,8502-8550,8555-8603,8608-8656,8661-8709,8714-8762,8767-8815,8820-8868,8873-8921,8926-8974,8979-9027,9032-9080,9085-9133,9138-9186,9191-9239,9244-9292,9297-9345,9350-9398,9403-9451,9456-9504,9509-9557,9562-9610,9615-9663,9668-9716,9721-9769,9774-9822,9827-9875,9880-9928,9933-9981,9986-10000,10005-10053,10058-10106,10111-10159,10164-10212,10217-10265,10270-10318,10323-10371,10376-10424,10429-10477,10482-10530,10535-10583,10588-10636,10641-10689,10694-10742,10747-10795,10800-10848,10853-10901,10906-10954,10959-11007,11012-11060,11065-11113,11118-11166,11171-11219,11224-11272,11277-11325,11330-11378,11383-11431,11436-11484,11489-11537,11542-11590,11595-11643,11648-11696,11701-11749,11754-11802,11807-11855,11860-11908,11913-11961,11966-12014,12019-12067,12072-12120,12125-12173,12178-12226,12231-12279,12284-12332,12337-12385,12390-12438,12443-12491,12496-12544,12549-12597,12602-12650,12655-12703,12708-12756,12761-12809,12814-12862,12867-12915,12920-12968,12973-13021,13026-13074,13079-13127,13132-13180,13185-13233,13238-13286,13291-13339,13344-13392,13397-13445,13450-13498,13503-13551,13556-13604,13609-13657,13662-13710,13715-13763,13768-13816,13821-13869,13874-13922,13927-13975,13980-14028,14033-14081,14086-14134,14139-14187,14192-14240,14245-14293,14298-14346,14351-14399,14404-14452,14457-14505,14510-14558,14563-14611,14616-14664,14669-14717,14722-14770,14775-14823,14828-14876,14881-14929,14934-14982,14987-15035,15040-15088,15093-15141,15146-15194,15199-15247,15252-15300,15305-15353,15358-15406,15411-15459,15464-15512,15517-15565,15570-15618,15623-15671,15676-15724,15729-15777,15782-15830,15835-15883,15888-15936,15941-15989,16000

## Query Match

Best Local Similarity 83.7%; Score 2055; DB 1; Length 1234;  
 Matches 355; Conservative 28; Mismatches 45; Indels 0; Gaps 0;

QY 1 EDDKPPRRNSLSSMSSEQLYSEGTQATYKCRPGYRLGTIVKCKNGEVPNPSR 60  
 |||||  
 Db 19 EDDKPPRRNSLSSMSSEQLYSEGTQATYKCRPGYRLGTIVKCKNGEVPNPSR 78  
 |||||  
 QY 61 ICRKPPGHDGDPFGSFRILAVGSEFEFGAKVYTCDEGTOLGELIDYRECDADGWTNDI 120  
 |||||  
 Db 79 ICRKPPGHDGDPFGSFRILAVGSEFEFGAKVYTCDEGTOLGELIDYRECDADGWTNDI 138  
 |||||  
 QY 121 PICEVVKCLPYLENGLRIVSGAAPEQYXFGQVRFECNNGSKIEGOKEMHCSEGLW 180  
 |||||  
 Db 139 PICEVVKCLPYLENGLRIVSGAAPEQYXFGQVRFECNNGSKIEGOKEMHCSEGLW 198  
 |||||  
 QY 181 SNKKPOCVETSLPPRENGDGIYKPYKENEFOYKCKGQFYKRRGAVCTGSSWNP 240  
 |||||  
 Db 199 SNKKPOCVETSLPPRENGDGIYKPYKENEFOYKCKGQFYKRRGAVCTGSSWNP 258  
 |||||  
 QY 241 QPSCSEMTCLTPYIPNGIYPRHRIKRIIDELIRECKNGEVPATRSVSCCTITGWPAP 300  
 |||||  
 Db 259 QPSCSEMTCLTPYIPNGIYPRHRIKRIIDELIRECKNGEVPATRSVSCCTITGWPAP 318  
 |||||  
 QY 301 RCLKPCDPFOFHRGLRYEESRRPYPVPIGKEYSYCCNGFTTSPQSTWYDIKCTVNG 360  
 |||||  
 Db 319 RCLKPCDPFOFHRGLRYEESRRPYPVPIGKEYSYCCNGFTTSPQSTWYDIKCTVNG 378  
 |||||  
 QY 361 WEPEVPLRCRIFHYEYEGSSYWMORRYIEGSAKYOCCHSGSYSPNODPYHCENWSP 420  
 |||||  
 Db 379 WEPEVPLRCRIFHYEYEGSSYWMORRYIEGSAKYOCCHSGSYSPNODPYHCENWSP 438  
 |||||

QY 421 PPKCVRIK 428  
 |||||  
 Db 439 PPKCIRIK 446

## RESULT 2

NBHRH  
 Complement factor H precursor, long splice form - human

C:Species: Homo sapiens (man)  
 C:Date: 31-Dec-1993 #sequence, revision 31-Dec-1993 #text, change 22-Jun-1999

C:Accession: S00254; A60238; A54726; A61565; A26505; I72654; S66298  
 R:Ripocher, J.; Day, A.J.; Harris, T.J.R.; Sim, R.B.

Biochem. J. 249, 593-602, 1988  
 A:Title: The complete amino acid sequence of human complement factor H.

A:Reference number: S00254; M01D:88134059  
 A:Accession: S00254

A:Molecule type: mRNA  
 A:Residues: 1-1231 <RIP>

A:Cross-references: EMBL:Y00716; NID:931964; PIDN:CAA68704.1; PID:931965  
 A:Note: 402-tyr was also found

A:Note: parts of this sequence, including the amino and carboxyl ends of the mature protein  
 R:Estallier, C.; Schwaible, W.; Dierich, M.; Weiss, E.H.

Eur. J. Immunol. 21, 799-802, 1991  
 A:Title: Human complement factor H: two factor H proteins are derived from alternatively

A:Reference number: A60238; M01D:91184292

A:Accession: A60238  
 A:Status: not compared with conceptual translation

A:Molecule type: mRNA  
 A:Residues: 1-56; 1177-1231 <EST>

A:Note: only portions of this 4.3 kilobase mRNA were sequenced  
 R:Day, A.J.; Ripocher, J.; Lyons, A.; McIntosh, B.; Harris, T.J.R.; Sim, R.B.

Biochem. J. 249, 593-602, 1988  
 A:Title: Sequence analysis of a cDNA clone encoding the C-terminal end of human complement factor H.

A:Reference number: A54726; M01D:88025472  
 A:Accession: A54726

A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA

A:Residues: 579-1231 <DAY>  
 A:Cross-references: GB:M17517; NID:9180497; PIDN:AAA52016.1; PID:9180498

A:Note: parts of this sequence were determined by protein sequencing  
 R:Ripocher, J.; Day, A.J.; Willis, A.C.; Belt, K.T.; Campbell, R.D.; Sim, R.B.

Biochem. J. 249, 593-602, 1988  
 A:Title: Partial characterization of human complement factor H by protein and cDNA

A:Reference number: A61565; M01D:86188123  
 A:Accession: A61565

A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA

A:Residues: 1050-1057, 'T', 1059-1102 <R12>  
 R:Sim, R.B.; Discipio, R.G.

Biochem. J. 205, 285-293, 1982  
 A:Title: Purification and structural studies on the complement-system control protein

A:Reference number: A26505; M01D:83048213  
 A:Accession: A26505

A:Molecule type: protein  
 A:Residues: 19-20, 'Q', 22-29, 'V', 31-33, 'Q', 35 <SIM>

R:Barlow, P.N.; Norman, D.G.; Steinkasserer, A.; Horne, T.J.; Pearce, J.; Driscoll, P.  
 Biochemistry 31, 3626-3634, 1992

A:Title: Solution structure of the fifth repeat of factor H: A second example of the  
 A:Reference number: A44551; M01D:92232649

A:Contents: annotation; NMR structure determination, residues 264-292  
 R:Norman, D.G.; Barlow, P.N.; Baron, M.; Day, A.J.; Sim, R.B.; Campbell, I.D.

J. Mol. Biol. 219, 717-725, 1991  
 A:Title: Three-dimensional structure of a complement control protein module in solution

A:Reference number: A49224; M01D:91278097  
 A:Contents: annotation; NMR structure determination, residues 927-985

R:Estallier, C.; Kolstien, V.; Schwaible, W.; Dierich, M.P.; Weiss, E.H.  
 J. Immunol. 146, 3190-3196, 1991

A:Title: Cloning of the 1.4-kb mRNA species of human complement factor H reveals a no

A:Reference number: I56100; M01D:91201892  
 A:Accession: I72654

A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA

A:Residues: 1047-1231 <RES>  
 A:Cross-references: GB:M65294; NID:9183766; PIDN:AAA35948.1; PID:9183767

R:Carroll, J.A.; Bates, R.C.; Smith, A.I.; Tetlow, T.; Arellano, A.; Gordon, D.L.; Burn  
 Biochim. Biophys. Acta 1289, 305-311, 1996

A:Title: Factor H co-purifies with thrombospondin isolated from platelet secretate.  
 A:Reference number: S66298; M01D:96205365

A:Accession: S66298  
 A:Status: preliminary

A:Molecule type: protein  
 A:Residues: 411-419, 574-578, 580-582 <CAR>

C:Comment: Factor H has also been found bound to cell membranes in an unknown manner.  
 C:Comment: Alternative transcripts of 4.3, 1.8, and 1.4 kilobases are expressed in 11v

C:Genetics: <HF1>  
 A:Gene: GDB:HF1; HF

A:Cross-references: GDB:120041; OMIM:134370  
 A:Map position: 1932-1932

C:Genetics: <HF2>  
 A:Gene: GDB:HF2; HF

A:Cross-references: GDB:129095  
 A:Map position: 1932-1932

A:Note: the correspondence between the two loci and the sequences indicated is unclear  
 A:Function:

A:Description: a cofactor in the inactivation of C3b by serine proteinase I; also involved in the alternative complement pathway

A:Pathway: complement alternate pathway

C:Superfamily: complement factor H; complement factor H repeat homology  
 C:Keywords: alternative splicing; complement alternate pathway; glycoprotein; plasma  
 F:1-18/Domain: signal sequence #status predicted <SIG>  
 F:19-1229/Product: complement factor H #status experimental <MP>  
 F:19-449/Product: complement factor H short splice form #status experimental <MP>  
 F:21-80/Domain: complement factor H repeat homology <FH01>  
 F:85-141/Domain: complement factor H repeat homology <FH02>  
 F:146-205/Domain: complement factor H repeat homology <FH03>  
 F:210-262/Domain: complement factor H repeat homology <FH04>  
 F:246-248/Region: cell attachment (R-G-D) motif  
 F:267-320/Domain: complement factor H repeat homology <FH05>  
 F:325-385/Domain: complement factor H repeat homology <FH06>  
 F:389-442/Domain: complement factor H repeat homology <FH07>  
 F:448-505/Domain: complement factor H repeat homology <FH08>  
 F:509-564/Domain: complement factor H repeat homology <FH09>  
 F:569-623/Domain: complement factor H repeat homology <FH10>  
 F:630-684/Domain: complement factor H repeat homology <FH11>  
 F:691-744/Domain: complement factor H repeat homology <FH12>  
 F:753-803/Domain: complement factor H repeat homology <FH13>  
 F:811-864/Domain: complement factor H repeat homology <FH14>  
 F:870-926/Domain: complement factor H repeat homology <FH15>  
 F:931-984/Domain: complement factor H repeat homology <FH16>  
 F:989-1043/Domain: complement factor H repeat homology <FH17>  
 F:1048-1102/Domain: complement factor H repeat homology <FH18>  
 F:1109-1163/Domain: complement factor H repeat homology <FH19>  
 F:1167-1228/Domain: complement factor H repeat homology <FH20>  
 F:21-66,52-80,85-129,114-141,146-192,178-205,210-251,237-262,267-309,294-320,325-374,357-1803,811-853,839-864,870-915,901-926,931-973,959-984,989-1032,1018-1043,1048-1091,1077-1529,802,822,882,911/Binding site: carbohydrate (Asn) (covalent) #status experimental  
 F:718,1029,1095/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 69.88; Score 1714; DB 1; Length 1231;  
 Best: Local Similarity 67.58; Pred. No. 4.5e-112;  
 Matches 289; Conservative 47; Mismatches 92; Indels 0; Gaps 0;

1 EDGCGPPPRESEITLGSMSSEOLYKCRPGYRTGTYKVKNGEWPNSNR 60  
 19 EDCEHLPFRRTTEITLGSMSDQTYPRGQYIKCRPGYRTGTYKVKNGEWPNSNR 78  
 61 ICRRPCGHPDTPFGSFRILAVGSEFEGAKVYVYTCDEYOLGELIDYRECDADGWTNDI 120  
 79 KCQKRPCHGPDTPFGFTLTGWNVEYGAAYTCNCGYLLGELINRRECDTGWTNDI 138  
 121 PICVAVCLPTLEENGRIVSGAEPDOEYFGGVNPRECSGKRTIEQAKHSGENGLW 180  
 139 PICVAVCLPTLAPENKIVSSAMEPREYHFGQAVRFVCSGKRIEEDDEMHCSDDGEFW 198  
 181 SNEPQCEVEISCLPPRENGDGYLKYKRENERFOYKCGKGFYKRGDVAVCTGSGMNP 240  
 199 SKERPKVEISCKSPDIYNGSPIOKITIKENEFQYKCNNGIYSENGDVAVCTGSGMNP 258  
 241 QPCEMTCLPTLYPINSIYTPHRIKRIIDIRYECNKGFPATRSVPYSCITIGWIPAP 300  
 259 LPSCSEKSDPRLYPNGDYSPLRIKHRTGDEITYCRNGFPATRGNAKCTSGMWIPAP 318  
 301 RCLSKPCDFPFKGRILYTESRRPYRPVPIGKREYSYCDNGFTTPSSYMDTLRCITYNG 360  
 319 RCLTKPCDYPRIKGGILYHEMRPRYPVAVGKYSYCDHFPSPSSYMDHITCTDGG 378  
 361 WEPEVPLKRCIFHYVEGESSYQWQRYIEGQSAKVOCHSGYSYLPNGDYYCTENGWSP 420  
 379 WSPAVPLKRCIFYPLNGYNGNNGRKFYCGSKIDVACHPQYALPKAOTYTTCENGMSP 438  
 421 PPKCVARIK 428  
 439 TPCIRIVK 446

RESULT 3  
 NBHHS  
 Complement factor H precursor, short splice form - human

N:Alternate names: complement factor H-related protein; complement protein H  
 C:Species: Homo sapiens (man)  
 C:Date: 31-Dec-1999 #sequence\_revision 23-Feb-1996 #text\_change 22-Jun-1999  
 C:Accession: S03013; B60238; A27877; A61103; A26505; S10479  
 R:Ripoche, J.; Day, A.J.; Harris, T.J.R.; Sim, R.B.  
 Biochem. J. 249, 593-602, 1988  
 A:Title: The complete amino acid sequence of human complement factor H.  
 A:Reference number: S00254; MUID:86134059  
 A:Accession: S03013  
 A:Molecule type: mRNA  
 A:Residues: 1-449 <RIP>  
 A:Cross-references: EMBL:X07523; EMBL:Y00716; NID:g32492; PIDN:CAA30403.1; PID:g75807  
 A:Note: Part of this sequence, including the amino end of the mature protein was conf  
 A:Note: 402-Tyr was also found  
 R:Estaller, C.; Schwaible, W.; Dierich, M.; Weiss, E.H.  
 Eur. J. Immunol. 21, 799-802, 1991  
 A:Title: Human complement factor H: two factor H proteins are derived from alternativ  
 A:Reference number: A60238; MUID:91184292  
 A:Accession: B60238  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-33,434-449 <EST>  
 A:Note: only portions of this 1.8 kilobase mRNA were sequenced  
 R:Schulz, T.F.; Schwaible, W.; Stanley, K.K.; Weiss, E.; Dierich, M.P.  
 Eur. J. Immunol. 16, 1351-1355, 1986  
 A:Title: Human complement factor H: isolation of cDNA clones and partial cDNA sequence  
 A:Reference number: A27877; MUID:87054207  
 A:Accession: A27877  
 A:Molecule type: mRNA  
 A:Residues: 1-17,55-401, 'Y', 403-449 <SCH>  
 A:Note: an additional nucleotide present within the codon for Glu-310 was thought to  
 R:Schwaible, W.; Zwirner, J.; Schulz, T.F.; Linke, R.P.; Dierich, M.P.; Weiss, E.H.  
 Eur. J. Immunol. 17, 1485-1489, 1987  
 A:Title: Human complement factor H: expression of an additional truncated gene produc  
 A:Reference number: A61103; MUID:88055295  
 A:Accession: A61103  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 19-20, 'Q', 22-29, 'V', 31-33, 'Q', 35 <SIM>  
 R:Barlow, P.N.; Norman, D.G.; Steinkasserer, A.; Horne, T.J.; Pearce, J.; Driscoll, P  
 Biochem. J. 31, 3626-3634, 1992  
 A:Title: Solution structure of the fifth repeat of factor H: A second example of the  
 A:Reference number: A44551; MUID:92232649  
 A:Contents: annotation; NMR structure determination, residues 264-292  
 R:Kristensen, T.; Weisdel, R.A.; Tack, B.F.  
 J. Immunol. 136, 3407-3411, 1986  
 A:Title: Structural analysis of human complement protein H: homology with C4b binding  
 A:Reference number: S10479; MUID:86169701  
 A:Accession: S10479  
 A:Molecule type: mRNA  
 A:Residues: 226-401, 'Y', 403-449 <KRI>  
 A:Cross-references: GB:M12383; NID:g180472; PIDN:AA52013.1; PID:g180473  
 C:Comment: Factor H has been found bound to cell membranes in an unknown manner.  
 C:Comment: Alternative transcripts of 4.3, 1.8, and 1.4 kilobases are expressed in liv  
 C:Genetics: <HP1>  
 A:Gene: GDB:HP1; HF  
 A:Cross-references: GDB:120041; OMIM:134370  
 A:Map position: 1q32-1q32  
 C:Genetics: <HP2>  
 A:Gene: GDB:HP2; HF  
 A:Cross-references: GDB:129095  
 A:Map position: 1q32-1q32  
 A:Note: the correspondence between the two loci and the sequences indicated is unclear  
 C:Function:  
 A:Description: a cofactor in the inactivation of C3b by serine proteinase I; also inc



```

QY      356  CTVMGEEVEVCLKQCIPHHYVEGESSYWRRTLYEGQSAKVCQCHSGSLPNQGDPTCYCTE 415
      ||| |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      135  CTAQGMEEVEVCVRKCVPHYVENGDSAWKEKTIYVGQSLKVCQCTNGSYSLQNGQDLMTC 194

QY      416  NGMSDPPKCVRIK 428
      ||| |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      195  NGMSPPPKCIRIK 207

RESULT  6
H35068
apolipoprotein H-related protein 23L1 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 27-Jul-1990 #sequence_revision 31-Dec-1993 #text_change 05-Nov-1999
C:Accession: H35068; A35069; B35069; I35068
R:VIK, D.P.; Munoz-Ganoves, P.; Kozono, H.; Martin, L.G.; Tack, B.F.; Chaplin, D.D.
J. Biol. Chem. 265, 3193-3201, 1990
A:Title: Identification and sequence analysis of four complement factor H-related trans-
A:Reference number: A35070; MUID:90153969
A:Accession: H35068
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-303 <VIK>
A:Cross-references: GB:M29007; NID:q192557; PIDN:AAA37413.1; PID:q309164; GB:J05259
A:Note: Translation of the nucleotide sequence is not complete
C:Superfamily: complement factor H repeat homology
C:Keywords: duplication
F:128-81/Domain: complement factor H repeat homology <FH01>
F:186-146/Domain: complement factor H repeat homology <FH02>
F:150-203/Domain: complement factor H repeat homology <FH03>
F:212-266/Domain: complement factor H repeat homology <FH04>

```

```

Query March 136: 33.0%; Score 811; DB 2; Length 303;
Best Local Similarity 70.8%; Pred. No. 1.1e-49;
Matches 136; Conservative 19; Mismatches 37; Indels 0; Gaps 0;

QY 236 SGWNPQSCIENTCLTPYIPNGIYTPHRIKRIIDDEIRYECKNGEYPATRSPVSKCTITG 295
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 15 TAWLSTANGKEETGCSPPYLNGIYTPHRIIHKSDDEIRIECKYGFYPAVYGSTVSKCTPTG 74

QY 236 WIPADPRLSKPCDPFQPFKNGRLIYEESRRPYFPVPIGKEYSYCDNGFTTSPQSQTWDYLR 355
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 75 WIPVPRCTLKPCFEPQFQYKGRLYEESLRPNPVPISIGNKYSYKCDNGFSPSGYSMDYLR 134

QY 356 CTYVNWMEPEVPLRCICITHIYVEGSSYWRKRIEGQSKAKVQCHSGISLPNQDPIYYCTE 415
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 135 CTAGQMEPEVPCPRVCRCFHYENGSDSTYWEKLYVQGSILKVCQYNGYSLQNGDITMTCPE 194

QY 416 NGWSPPKCVRI 427
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 195 NGWSPPKCIKI 206

RESULT 7
S46199
probable complement regulatory plasma protein SBI - barred sand bass
C:Species: Parablax neblifer (barred sand bass)
C:date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 08-Oct-1999
C:Accession: S46199; S77894
R:Daumen, A., Kaideh, T., Zipfel, P.F., Gligl, I.
Biochem. J. 301, 391-397, 1994
A:Title: Cloning and characterization of a cDNA representing a putative complement-regu
A:Reference number: S46199; MUID:94318039
A:Accession: S46199
A:Molecule type: mRNA
A:Residues: 1-1053 <DNA1>
A:Cross-References: EMBL:121703; NID:9639894; PIDN:AAA92556.1; PID:9639895
A:Experimental source: liver
A:Accession: S77894
A:Molecule type: protein
A:Residues: 526-532, 'X', 534-537, 809-817, 'X', 819-826 <DNA2>
C:Genetics:

```

A:Gene: SB1  
C:Superfamily: complement factor H repeat homology  
C:Keywords: glycoprotein  
E:89-145/Domain: complement factor H repeat homology <FH01>  
E:334-389/Domain: complement factor H repeat homology <FH02>  
E:450-502/Domain: complement factor H repeat homology <FH03>  
E:569-624/Domain: complement factor H repeat homology <FH04>  
E:682-738/Domain: complement factor H repeat homology <FH05>  
E:743-802/Domain: complement factor H repeat homology <FH06>  
E:935-989/Domain: complement factor H repeat homology <FH07>  
E:993-1052/Domain: complement factor H repeat homology <FH08>

Query March 23.2% Score 569.5; Db 2; Length 1053;  
Best Local Similarity 26.7% Pred. No. 3,1e-33;  
Matches 148; Conservative 60; Mismatches 169; Indels 177; Gaps 19;

```
QY      21 EQLYSEGTQATYKCRPGYRLLGTLTKVCKNGEWPSPNSRICKRRCCHPGDTPPGSFRLL 80
Db      47 EASYPGGQROVAVGCVGVS--GFEFLVCVECKMETRGAK--COPRSCGHPDADQAFAPFHL 102
QY      81 AVGSFEFGARVYVTCDSGYOLLGEIDYREDDAQMTNDIPICEVYKCLPATELENGRIY 140
Db      103 AEGNDFVGSKVYVTCQGYQOMSRINRCVABGMDGVVPCESQOC-PLIHVDNNYOV 161
QY      141 SGAEPDDEYFVGCVAFRECNSGFKI-EGQKEMHCSENGLMSNEKPOCEVIEISCLPVRVY 199
Db      162 IGGPE---EALFGVNVYRSCKSRSRELLDGSPELYCDEKMDSGVPYKPKATTCALPIREN 218
QY      200 GDGIYLFKPYKENERFOYKCKQGFYVKERGAVCTGSG---WNPQSCSEMTCL----- 250
Db      219 GNVPGAIRREYKENDVLHAYECDRAEFKHIDR-PSCTIKQIKAEWSPTPLCESIKCRLTIMD 277
QY      251 -TPYLP-----NGIYT----- 260
Db      278 GTRREPAYRNLFSPEGTLKVICARTSWISTPOEISVYTTCCODNGEWSIRPPOCEVYRCSNR 337
QY      261 -PHRI-----KHRIDDEIRCECKNGFVPATRSPPVSKCTITGMIPAP----- 300
Db      338 RPEHNSDMDVSMWEVETLDDNTRVWCXKRGYKRTGVTWATCGRNMGMNPNPLCEVYKTSKE 397
QY      301 -----RCSLKPDPDFPQPKH 314
Db      398 NIQDAVIVGTDKQIYNLNQKAIYACGEGNRRKITLTCGECNGMSGDRKCTVAPRCPPLPRDP 457
QY      315 GRLYVEESRRPFPYPICKEYSYCDNGEFTTPSQSYWYLYRCTGVNGMPEVPCLROCI-- 372
Db      458 NGFF---RGPI---TGRVLYYTCCKDGKLFLEGMMAEAKCYDGVW---PELTTCISLN 505
QY      373 -----FHYVEGESSYWMORYLEGOSAKVVOCHSGYSLBNPDQDTYYCTENGW----- 418
Db      506 TTGCGKFPFLPNAEVI---RRYPEVQTVQYICNGYS--TQANSFCSENGNMLLYGLSPDQ 560
QY      419 -----SPPPK 423
Db      561 ICTLRADYCGPPE 574

RESULT      8
D35069      complement factor H-related protein 9c4 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 27-Jul-1990 #sequence_revision 31-Dec-1993 #text_change 20-Aug-1999
C:Accession: A35070; B35070; C35070; D35070; E35069; F35070;
J:ViR. D.P.: Munoz-Caroves, P.; Kozono, H.; Martin, L.G.; Tack, B.F.; Chaplin, D.D.
J: Biol. Chem. 265, 3193-3201, 1990
A:Title: Identification and sequence analysis of four complement factor H-related tra
A:Reference number: A35070; MUID:90153969
A:Accession: D35069
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-808 <VIR>
A:Cross-references: GB:M29009; NID:g192563; PIDN:AAA37416.1; PID:g309166; GB:J05259
```



[illegible]

Db 103 YKFCVKKRCNGPELLNG--QVIVKTDFSGFSEIEFSCSEGYVLISANSYCOLQDKV 160  
QY 116 -WTNDIPICEVVKCLPYTELENGRIYVGAAPDOERY-FGQVAFECNSGFKIEGQKEM 173  
Db 161 VMSDPLPOCIACKEPPTISNGRHNG--DEDFYTSSTVYSCDDRFPSMLGKASIS 216  
QY 174 CS-EN---GLMSNEKQCVESICLPPRVENG---DGIYLPVKYKEMRFRQKCKGQFVYK 226  
Db 217 CRVENNTGVWSSPPSCKKVICQVPPVKDKITSG--FGPIYTYOQSIYACNKG--R 272  
QY 227 ERGDVAV--C-TGSGMN-POPSCEMTCL-TPYIPNGIYTPHRIKHRDDE-----1 272  
Db 273 LEBDSILHCEADSNMNPPTCTELNGLGIPHPHAI--ERDHOQTEQOQYVDIGFVL 330  
QY 273 RECKNGFFPAIRSPVSKCTT-----GWIPARCSLKPCDFPOF-KHG--RLYERSRR 324  
Db 331 SYKCHGKYRETDGPT---TVTCOSNLEMSPYTECKEVCCPEPNLNNYSGITLHRRSTS 367  
QY 325 PYEPVPIKEYSYCDNG-----F 343  
Db 368 THCTYISGDIISTECHKYMPDALCTKHGTSWSPRTEPCRDCKSPPIAHGKHVYSKFP 447  
QY 344 TPFSQSYWD-----YLRCYVNGWEPEVP-CLROCIFFHYVERGESSYMQRRYIEG 391  
Db 448 TRDHQAVYECODKGYLLVGAKEILSCTSSGMSPANPOCKALCLKEIRGLSVKRVYEP 507  
QY 392 QSAKVCHSGYSLPNCQDPTYYCTEN-GWSBP-PKC 424  
Db 508 EITTIQCESGYSV-GSENIITCSEDRTWYEPVKC 541

## RESULT 12

A45222

Complement factor H-related protein DOWN16 precursor - human  
C:Species: Homo sapiens (man)  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 31-Mar-2000  
C:Accession: A45222; S29609

R:Skerka, C.; Kuhn, S.; Gunther, K.; Lingelbach, K.; Zlpfel, P.F.  
J. Biol. Chem. 268, 2904-2908, 1993  
A:Title: A novel short consensus repeat-containing molecule is related to human compleme

A:Reference number: A45222; MUID:93155112  
A:Accession: A45222

A:Molecule type: mRNA  
A:Residues: 1-331 <SKD>

A:Cross-references: EMBL:X68679; NID:g30869; PID:g30870  
A:Experimental source: liver

A:Note: sequence extracted from NCBI backbone (NCBIP:124269)  
C:Genetics:

A:Gene: GDB:HFL4  
A:Cross-references: GDB:137201

C:Superfamily: complement C3b/Epstein-Barr virus receptor; complement factor H repeat hc  
C:Keywords: glycoprotein

F:1-18/Domain: signal sequence #status predicted <SIG>  
F:19-331/Product: complement factor H-related protein DOWN16 #status predicted <MAT>

F:23-83/Domain: complement factor H repeat homology <FH01>  
F:87-141/Domain: complement factor H repeat homology <FH02>

F:147-204/Domain: complement factor H repeat homology <FH03>  
F:211-265/Domain: complement factor H repeat homology <FH04>

F:269-330/Domain: complement factor H repeat homology <FH05>  
F:108,186,206,310/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 19.0%; Score 466.5; DB 2; Length 331;  
Best Local Similarity 56.3%; Pred. No. 1.3e-25;

Matches 76; Conservative 18; Mismatches 40; Indels 1; Gaps 1;

QY 293 ITGMIAPRCSLKPCDFPOFKHGRLYVESRRPYPIGKEYSYCDNGFTPSQSYWD 352  
Db 9 LTLWVSCANGVKKPCDFDIKHGGLFHEHMMRRPYPAVAKYISYCDHEHFTPEGSGSYWD 68

QY 353 YLRCTVNGWEVVCCLRCIFHYVEYEGSSYQWRKRIYQSAKVCHSGYSLP-NDPT 411  
Db 69 YIHCTQMSWSPAVBCLRCYFPYLENGYNYGRKFGVGNSTEVACHDGYGLPKVYQTTV 128

QY 412 YCTENGWSPPKCYR 426  
Db 129 TCTENGWSPPTPCIR 143

## RESULT 13

A45900

Complement C3b receptor type 2 long form precursor - mouse (fragment)

C:Species: Mus musculus (house mouse)  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 08-Oct-1999  
C:Accession: A45900; I48306

R:Kurtz, C.B.; O'Toole, E.; Christensen, S.M.; Weis, J.H.  
J. Immunol. 144, 3581-3591, 1990

A:Title: The murine complement receptor gene family. IV. Alternative splicing of Cr2  
A:Reference number: A45900; MUID:90229754

A:Accession: A45900

A:Status: preliminary  
A:Molecule type: mRNA

A:Residues: 1-676 <KUR>  
A:Cross-references: GB:M36470

A:Experimental source: clone 31-1  
R:Kim, Y.U.; Kinoshita, T.; Molina, H.; Hourcade, D.; Seya, T.; Wagner, L.W.; Holers,

J. Exp. Med. 181, 151-159, 1995

A:Title: Mouse complement regulatory protein Cr1/p65 uses the specific mechanisms of  
A:Reference number: I48306; MUID:95105691

A:Accession: I48306

A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA

A:Residues: 21-367 <RES>  
A:Cross-references: EMBL:U17128; NID:g595980; PID:AAA78271.1; PID:g595982

C:Genetics:

A:Gene: Cr2  
A:Introns: 80/1; 113/2; 142/1; 275/1; 333/1

C:Superfamily: complement factor H repeat homology  
F:22-78/Domain: complement factor H repeat homology

F:83-140/Domain: complement factor H repeat homology <FH01>  
F:145-211/Domain: complement factor H repeat homology <FH02>

F:217-272/Domain: complement factor H repeat homology <FH03>  
F:276-331/Domain: complement factor H repeat homology <FH04>

F:336-394/Domain: complement factor H repeat homology <FH05>  
F:399-458/Domain: complement factor H repeat homology <FH06>

F:467-523/Domain: complement factor H repeat homology <FH07>  
F:531-587/Domain: complement factor H repeat homology <FH08>

F:592-648/Domain: complement factor H repeat homology <FH09>  
F:592-648/Domain: complement factor H repeat homology <FH10>

Query Match 18.5%; Score 453; DB 2; Length 676;  
Best Local Similarity 28.6%; Pred. No. 2.6e-24;

Matches 126; Conservative 73; Mismatches 178; Indels 64; Gaps 27;

QY 24 YSEGTQATYKCRPGYRTTGTIVKCKNGEWPNSRICKRRCGHPGPTPGSPFLAVG 83  
Db 42 FLIGTITWEKCKRPGYFRKRSFIITCLETSKW--SDAQCFKRRPCANPQPLGSHVING 99

QY 84 SEFPEGAKVYVYCDGSIYQLGEIDYRECDAG--WTNDIPICEVVKCLPYTELENGRIY 140  
Db 100 --IEFGSTIYSCNOGYRLIGDSATCIYSDNTVMWMDNMPCEISPCSPAISNGDFY 157

QY 141 SGAAPDDEYVGGVAFECNSG-----FKIEGQKEMHCSN-----GLMSNEKQCV-E 189  
Db 158 SSSRD---SFFYGMVYTYCHTGKNRREKLDLVGKRSYICYSKDQVGIWNSPPQCIPT 214

QY 190 ISCLPPRVENG--DGIYLPVKYKEMRFRQKCKGQFVYKKEGDAVC-TGSGMNPO-PSC 244  
Db 215 VACPMPREIENGLESV--FKHSFPLNDYIVYFKCKSGFTMKSGRIAMQPNKMSWSPPLPTC 272

QY 245 EEMTCLTY-IPNGIYTHRIKHRIDDEIRIECKNGFFPAIRSPVSKCTTIG-WI-PAIR 301  
Db 273 -FMGLCPQONILHGDYNNKDEFFSVGQKVSYYCNPQ-YTLIGTINLVECTSLGTMTNPT 330

QY 302 CSLKPCD-FP-QFKIGRLYVESRRPYPIGKEYSYCDNGFT---TPSQSYWDYLR 356  
Db 302 CSLKPCD-FP-QFKIGRLYVESRRPYPIGKEYSYCDNGFT---TPSQSYWDYLR 356



Tue Nov 21 16:57:31 2000

us-09-316-163-14.rpr

Page 10

[illegible]

Search completed: November 21, 2000, 16:47:45  
Job time: 374 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2000, 16:49:24 ; Search time 27.94 seconds  
(without alignments)  
489.396 Million cell updates/sec

Title: US-09-316-163-14  
Perfect score: 2454  
Sequence: 1 EDCKGPPRENSSEILSGSMS.....DTYCTENGWSPPCVRIK 428

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 87993 seqs, 31947931 residues  
Total number of hits satisfying chosen parameters: 87993

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2055	83.7	1234	CFAH_MOUSE	P06909 mus musculus
2	1714	69.8	1231	CFAH_HUMAN	P08603 homo sapien
3	481.5	19.6	597	C4BP_HUMAN	P04003 homo sapien
4	474.5	19.3	610	C4BP_BOVIN	Q28065 bos taurus
5	466.5	19.0	331	CFHD_HUMAN	Q02985 homo sapien
6	449	18.3	558	C4BP_RAT	Q63514
7	437	17.8	2039	CRL_HUMAN	P17927 homo sapien
8	384.5	15.7	469	C4BP_MOUSE	P08607 mus musculus
9	373	15.2	1025	CR2_MOUSE	P19070 mus musculus
10	368	15.0	1033	CR2_HUMAN	P20023 homo sapien
11	336	13.7	661	F13B_HUMAN	P05160 homo sapien
12	332	13.5	668	F13B_MOUSE	Q07968 mus musculus
13	335	12.8	330	CFHL_HUMAN	Q03591 homo sapien
14	314.5	12.8	263	VCP_VACCV	P10398 vaccinia vl
15	310.5	12.7	768	LEM3_MOUSE	Q01102 mus musculus
16	310	12.6	270	CFHE_HUMAN	P36980 homo sapien
17	309.5	12.6	830	LEM3_HUMAN	P16109 homo sapien
18	298.5	12.2	377	MCP_HUMAN	P15529 homo sapien
19	296	12.1	768	LEM3_RAT	P98106 ratu
20	290	11.8	610	LEM2_HUMAN	P16581 homo sapien
21	282.5	11.5	769	LEM3_SHEEP	P98109 ov
22	277.5	11.3	611	LEM2_CANFA	P00690 mus musculus
23	272	11.1	612	LEM2_MOUSE	Q01339 mus musculus
24	271.5	11.1	345	APDH_MOUSE	P02749 homo sapien
25	267.5	10.9	345	APDH_HUMAN	O60401 cavia porce
26	267.5	10.7	507	DAF_MOUSE	P33703 canis famill
27	262.5	10.5	345	APDH_CANFA	P27113 oryctolagus
28	256.5	10.3	551	LEM2_RABIT	P98110 sus scrofa
29	255	10.4	360	CCPH_HSVSA	P42201 bos taurus
30	246.5	10.0	484	LEM2_PIG	P08174 homo sapien
31	246.5	10.0	646	LEM3_BOVIN	Q61475 mus musculus
32	244.5	9.9	381	DAF_HUMAN	
33	243		390	DAF_MOUSE	

## ALIGNMENTS

RESULT 1	CFAH_MOUSE	STANDARD:	PRT: 1234 AA.	HIG_DROME	Q09101 drosophila
AC	P06909;			340	P49457 pongo pygma
DT	01-JAN-1988 (Rel. 06, Created)			345	P17690 bos taurus
DT	01-JAN-1988 (Rel. 06, Last sequence update)			345	P17690 bos taurus
DT	01-NOV-1997 (Rel. 35, Last annotation update)			345	P17690 bos taurus
DE	COMPLEMENT FACTOR H PRECURSOR (PROTEIN BETA-1-H).			345	P17690 bos taurus
GN	CFH.			345	P17690 bos taurus
OS	Mus musculus (Mouse).			345	P17690 bos taurus
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			345	P17690 bos taurus
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			345	P17690 bos taurus
RN	[1]			345	P17690 bos taurus
RP	SEQUENCE FROM N.A.			345	P17690 bos taurus
RX	MEDLINE: 86233353.			345	P17690 bos taurus
RA	Kristensen T., Tack B.F., Vik D.P.;			345	P17690 bos taurus
RT	"Murine protein H is comprised of 20 repeating units, 61 amino acids in length."			345	P17690 bos taurus
RL	Proc. Natl. Acad. Sci. U.S.A. 83:3963-3967(1986).			345	P17690 bos taurus
RN	[2]			345	P17690 bos taurus
RP	SEQUENCE OF 1-19 FROM N.A.			345	P17690 bos taurus
RX	STRAIN-BALB/C.			345	P17690 bos taurus
RA	MEDLINE: 90148935.			345	P17690 bos taurus
RT	Munoz-Canoves P., Tack B.F., Vik D.P.;			345	P17690 bos taurus
RL	"Analysis of complement factor H mRNA expression: dexamethasone and IFN-gamma increase the level of H in L cells."			345	P17690 bos taurus
RN	Biochemistry 28:9891-9897(1989).			345	P17690 bos taurus
RP	[3]			345	P17690 bos taurus
RX	SEQUENCE OF 1-18 FROM N.A.			345	P17690 bos taurus
RA	MEDLINE: 90111033.			345	P17690 bos taurus
RT	Natsune-Sakai S., Nonaka M., Harada Y.N., Shreffler D.C., Moriaki K.;			345	P17690 bos taurus
RL	"Demonstration of an unusual allelic variation of mouse factor H by the complete cDNA sequence of the H.2 allotype."			345	P17690 bos taurus
RP	J. Immunol. 144:358-362(1990).			345	P17690 bos taurus
RX	- FUNCTION: FACTOR H FUNCTIONS AS A COFACTOR IN THE INACTIVATION OF C3B BY FACTOR I AND ALSO INCREASES THE RATE OF DISSOCIATION OF THE C3BB COMPLEX (C3 CONVERTASE) AND THE (C3)NBB COMPLEX (C5 CONVERTASE) IN THE ALTERNATIVE COMPLEMENT PATHWAY.			345	P17690 bos taurus
RA	- POLYMORPHISM: TWO CODOMINANT ALLELES OF FACTOR H ARE PRESENT IN MICE.			345	P17690 bos taurus
RT	- SIMILARITY: CONTAINS 20 SUSHI (SCR) REPEATS.			345	P17690 bos taurus
RL	- This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb.ch/announce/ or send an email to license@isb-sib.ch).			345	P17690 bos taurus
RP	EMBL: M12660; AAA37759.1; -			345	P17690 bos taurus
RX	EMBL: J02891; AAA37759.1; -			345	P17690 bos taurus
RA	EMBL: M31979; AAA37762.1; -			345	P17690 bos taurus
RT	PIR: A26154; NEMSH.			345	P17690 bos taurus
RL	HSSP: P08603; IHFI.			345	P17690 bos taurus



RT "Structural analysis of human complement protein H: homology with C4b  
RT binding protein, beta 2-glycoprotein I, and the Ba fragment of B2.";  
RL J. Immunol. 136:3407-3411(1986).  
[4]  
RP SEQUENCE OF 1047-1231 FROM N.A.  
RX MEDLINE: 91201892.  
RA Estalier C., Koistinen V., Schaeble W., Dierich M.P., Weiss E.H.;  
RT "Cloning of the 1.4-kb mRNA species of human complement factor H."  
RT reveals a novel member of the short consensus repeat family related  
RT to the carboxy terminal of the classical 150-kDa molecule.<sup>1</sup>  
RL J. Immunol. 146:3190-3196(1991).  
[5]  
RP SEQUENCE OF 19-35.  
RX MEDLINE: 83048213.  
RA Sim R.B., Discipio R.G.;  
RT "Purification and structural studies on the complement-system control  
RT protein beta 1H (factor H).";  
RL Biochem. J. 205:285-293(1982).  
[6]  
RP STRUCTURE BY NMR OF 927-985 (SUSHI 16).  
RX MEDLINE: 91278097.  
RA Norman D.G., Barlow P.N., Baron M., Day A.J., Sim B., Campbell I.D.;  
RT "Three-dimensional structure of a complement control protein module  
RT in solution.";  
RL J. Mol. Biol. 219:717-725(1991).  
[7]  
RP STRUCTURE BY NMR OF 264-322 (SUSHI 5).  
RX MEDLINE: 92232649.  
RA Barlow P.N., Norman D.G., Steinkasserer A., Horne T.J., Pearce J.,  
RL Driscoll P.C., Sim B., Campbell I.D.;  
RT "Solution structure of the fifth repeat of factor H: a second example  
RT of the complement control protein module.";  
RL Biochemistry 31:3626-3634(1992).  
[8]  
RP STRUCTURE BY NMR OF 866-985 (SUSHIS 15 AND 16).  
RX MEDLINE: 93323119.  
RA Barlow P.N., Steinkasserer A., Norman D.G., Kieffer B., Wiles A.P.,  
RL Sim B., Campbell I.D.;  
RT "Solution structure of a pair of complement modules by nuclear  
RT magnetic resonance.";  
RL J. Mol. Biol. 232:268-284(1993).  
[9]  
RP FUNCTION: FACTOR H FUNCTIONS AS A COFACTOR IN THE INACTIVATION OF  
CC C3B BY FACTOR I AND ALSO INCREASES THE RATE OF DISSOCIATION OF THE  
CC C3BB COMPLEX (C3 CONVERTASE) AND THE (C3b)NMB COMPLEX (C5  
CC CONVERTASE) IN THE ALTERNATIVE COMPLEMENT PATHWAY.  
CC -1- SIMILARITY: CONTAINS 20 SUSHI (SCR) REPEATS.  
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION  
CC 341 ONWARD DUE TO A FRAMESHIFT.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC  
DR EMBL: Y00716; CAA68704.1; -  
DR EMBL: X04697; CAB41739.1; ALT\_FRAME.  
DR EMBL: M65294; AAA35948.1; -  
DR PIR: S00254; NBH0H.  
DR PIR: S00254; S00254.  
DR PIR: S03013; S03013.  
DR PDB: 1HCC; 15-APR-92.  
DR PDB: 1HHY; 15-JUL-93.  
DR PDB: 1HHI; 15-JUL-93.  
DR MIM: 134370; -  
DR INTERPRO: IPR000436; -  
DR PFAM: PF00084; sushi; 20.  
KW Complement alternate pathway; Plasma; Glycoprotein; Repeat; Sushi;  
KW Signal; 3D-structure; Polymorphism.  
FT SIGNAL 1 18  
FT CHAIN 19 1231 COMPLEMENT FACTOR H.

FT	DOMAIN	20	1229	20 X SUSHI (SCR) REPEATS.
FT	REPEAT	20	81	SUSHI 1.
FT	REPEAT	84	142	SUSHI 2.
FT	REPEAT	145	206	SUSHI 3.
FT	REPEAT	209	263	SUSHI 4.
FT	REPEAT	266	321	SUSHI 5.
FT	REPEAT	324	386	SUSHI 6.
FT	REPEAT	388	443	SUSHI 7.
FT	REPEAT	447	506	SUSHI 8.
FT	REPEAT	508	565	SUSHI 9.
FT	REPEAT	568	624	SUSHI 10.
FT	REPEAT	629	685	SUSHI 11.
FT	REPEAT	690	745	SUSHI 12.
FT	REPEAT	752	804	SUSHI 13.
FT	REPEAT	810	865	SUSHI 14.
FT	REPEAT	869	927	SUSHI 15.
FT	REPEAT	930	985	SUSHI 16.
FT	REPEAT	988	1044	SUSHI 17.
FT	REPEAT	1047	1103	SUSHI 18.
FT	REPEAT	1108	1164	SUSHI 19.
FT	REPEAT	1166	1229	SUSHI 20.
FT	DISULFID	21	66	BY SIMILARITY.
FT	DISULFID	52	80	BY SIMILARITY.
FT	DISULFID	85	129	BY SIMILARITY.
FT	DISULFID	114	141	BY SIMILARITY.
FT	DISULFID	146	192	BY SIMILARITY.
FT	DISULFID	178	205	BY SIMILARITY.
FT	DISULFID	210	251	BY SIMILARITY.
FT	DISULFID	237	262	BY SIMILARITY.
FT	DISULFID	267	309	BY SIMILARITY.
FT	DISULFID	294	320	BY SIMILARITY.
FT	DISULFID	325	374	BY SIMILARITY.
FT	DISULFID	357	385	BY SIMILARITY.
FT	DISULFID	389	431	BY SIMILARITY.
FT	DISULFID	416	442	BY SIMILARITY.
FT	DISULFID	448	494	BY SIMILARITY.
FT	DISULFID	477	505	BY SIMILARITY.
FT	DISULFID	509	553	BY SIMILARITY.
FT	DISULFID	536	564	BY SIMILARITY.
FT	DISULFID	569	611	BY SIMILARITY.
FT	DISULFID	597	623	BY SIMILARITY.
FT	DISULFID	630	673	BY SIMILARITY.
FT	DISULFID	659	684	BY SIMILARITY.
FT	DISULFID	691	733	BY SIMILARITY.
FT	DISULFID	719	744	BY SIMILARITY.
FT	DISULFID	753	792	BY SIMILARITY.
FT	DISULFID	781	803	BY SIMILARITY.
FT	DISULFID	811	853	BY SIMILARITY.
FT	DISULFID	839	864	BY SIMILARITY.
FT	DISULFID	870	915	BY SIMILARITY.
FT	DISULFID	901	926	BY SIMILARITY.
FT	DISULFID	931	973	BY SIMILARITY.
FT	DISULFID	959	984	BY SIMILARITY.
FT	DISULFID	989	1032	BY SIMILARITY.
FT	DISULFID	1018	1043	BY SIMILARITY.
FT	DISULFID	1048	1091	BY SIMILARITY.
FT	DISULFID	1077	1102	BY SIMILARITY.
FT	DISULFID	1109	1152	BY SIMILARITY.
FT	DISULFID	1138	1163	BY SIMILARITY.
FT	DISULFID	1167	1218	BY SIMILARITY.
FT	DISULFID	1201	1228	BY SIMILARITY.
FT	CARBOHYD	529	529	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	718	718	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	802	802	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	822	822	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	882	882	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	911	911	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1029	1029	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1095	1095	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARIANT	402	402	H -> Y.
FT	CONFLICT	21	21	/FTID=VAR_001979.
FT	CONFLICT	30	30	C -> Q (IN REF. 3).
FT	CONFLICT			T -> V (IN REF. 3).

FT	CONFLICT	34	34	T -> Q (IN REF. 3).
FT	CONFLICT	53	54	RP -> IL (IN REF. 2).
FT	STRAND	870	871	
FT	STRAND	876	876	
FT	TURN	877	878	
FT	STRAND	879	881	
FT	STRAND	890	891	
FT	TURN	893	894	
FT	STRAND	896	900	
FT	TURN	903	904	
FT	STRAND	907	907	
FT	STRAND	912	916	
FT	TURN	917	918	
FT	STRAND	919	920	
FT	STRAND	926	926	
SO	SEQUENCE	1231 AA;	139125 MM;	C65EC8CF8800B3FD CRC64;

Query Match	69.8%;	Score 1714;	DB 1;	Length 1231;
Best Local Similarity	67.5%;	Pred. No. 6.4e-123;		
Matches 289;	Conservative 47;	Mismatches 92;	Indels 0;	Gaps 0;

Matches 289; Conservative 47; Mismatches 92; Indels 0; Gaps 0;

QY	1	EDCGPPRENSSEILGSWSBOLSECTOATYKCRPGYRLGTIVKCKNGEWNVPSNPR	60
Db	19	EDCNEILPPRRNTTELITGSWSQDITPEGTQATYKCRPGYRSLGNVIMVCRGGEWALNPLR	78
QY	61	ICRRPCGHPGDPFGSFRILAVSGSEEFEGAKVYTCDEGYOLLGELIDYRECDADGWTNDI	120
Db	79	KCKRRPCGHPDPFGFTLLGAVFEGYAVVATYCNESGYOLLGEINYPRECDPDGWTNDI	138
QY	121	PICEVVKCLPTLELNGRIYSGAEPDOEYFGVAVFECNNSGKTEGQCEMHCSENGIM	180
Db	139	PICEVVKCLPTLADENGRIVYSSAEPREHFGAVFVCONSGKITGDEMHCSODGEW	198
QY	181	SNEKPOCEISETLPPRYENGIGIYLKDVYKENERFOYKCKOGFYEKKERGAVCTGSGMP	240
Db	199	SKERPKCEISKSPDIYVINGSPISQKILYKENERFOYKCKMNGYEYSRGGAVCTESGMR	258
QY	241	QPSCEBMTCLPTIYPNGIYTPHRIKXHNIDEIRYECKNGFYPATRSVYSKCTITGWIPAP	300
Db	259	LPSEEEKSCDMPIYIPNDISPLRIKXHTGDEITVQCKNGFYPATRGWTACTGTGWTIPAP	318
QY	301	RCSLCKPCEPFQFKHGRLYEESRRPYFVPVPIGKFSYCYCNGGFTTSPQSNWDLIRCTVNG	360
Db	319	RCTLAKPCDIYPIKIKGGILXHEMRKRPYPVAVGKITYSTYCOEHFETTPSGSTWMDIHCTODG	378
QY	361	WEPEVPLCQICFIHYVEGESSYVQRRYIEGQSAKVQCHSGYSLIPNGQDTYYCTENGSWP	420
Db	379	WSPAVPCLRKCYFPYLENGYQUNHGRKRFVQGSIDVACHGYALPCKQTIVTCTENGSWP	438
QY	421	PPKCYRIK	428
Db	439	TPRCIRVX	446

RESULT	3
CABP_HUMAN	
ID	
AC	P04003
DT	23-OCT-1986 (Rel. 02, Created)
DT	01-OCT-1993 (Rel. 27, Last sequence update)
DT	01-NOV-1997 (Rel. 35, Last annotation update)
DE	CAB-BINDING PROTEIN ALPHA CHAIN PRECURSOR (PROLINE-RICH PROTEIN) (PRP).
GN	CABPA OR CABP.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN	[1]
RP	SEQUENCE FROM N.A.
RX	TISSUE=LIVER;
RX	MEDLINE: 90073699.
RA	Matsuguchi T., Okamura S., Aso T., Sata T., Niho Y.,

RT "Molecular cloning of the cDNA coding for prolins-rich protein (PRP):  
 RT identity of PRP as C4b-binding protein.";  
 RL Biochem. Biophys. Res. Commun. 105:138-144(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE; 9113139.  
 RX Aso T., Okamura S., Matsuguchi T., Sakamoto N., Sata T., Nihio Y.;  
 RT "Genomic organization of the alpha chain of the human C4b-binding  
 RL protein gene.";  
 RN Biochem. Biophys. Res. Commun. 174:222-227(1991).  
 RP [3]  
 RA SEQUENCE OF 9-81 FROM N.A.  
 RX MEDLINE; 88242821.  
 RL Lintin S.J., Lewin A.R., Reid K.B.M.;  
 RT "Derivation of the sequence of the signal peptide in human  
 RL C4b-binding protein and interspecies cross-hybridisation of the C4bp  
 RL cDNA sequence.";  
 RN FEBS Lett. 232:328-332(1988).  
 RP [4]  
 RA SEQUENCE OF 203-288 FROM N.A.  
 RX MEDLINE; 86301119.  
 RL Lintin S.J., Reid K.B.M.;  
 RT "Studies on the structure of the human C4b-binding protein gene.";  
 RN FEBS Lett. 204:77-81(1986).  
 RP [5]  
 RA SEQUENCE OF 80-597 FROM N.A.  
 RX MEDLINE; 86025405.  
 RL Chung L.P., Bentley D.R., Reid K.B.M.;  
 RT "Molecular cloning and characterization of the cDNA coding for C4b-  
 RL binding protein, a regulatory protein of the classical pathway of the  
 RL human complement system.";  
 RN Biochem. J. 230:133-141(1985).  
 RP [6]  
 RA SEQUENCE OF 49-88.  
 RX MEDLINE; 85296001.  
 RL Chung L.P., Gagnon J., Reid K.B.M.;  
 RT "Antico acid sequence studies of human C4b-binding protein: N-terminal  
 RL sequence analysis and alignment of the fragments produced by limited  
 RL proteolysis with chymotrypsin and the peptides produced by cyanogen  
 RL bromide treatment.";  
 RN Mol. Immunol. 22:427-435(1985).  
 RP [7]  
 RA ELECTRON MICROSCOPY, 3-DIMENSIONAL STRUCTURE, AND LIGAND BINDING.  
 RX MEDLINE; 83221615.  
 RL Dahlback B., Smith C.A., Mueller-Eberhard H.J.;  
 RT "Visualization of human C4b-binding protein and its complexes with  
 RL vitamin K-dependent protein S and complement protein C4b.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 80:3461-3465(1983).  
 CC -1- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT  
 CC ACTIVATION. IT BINDS AS A COPACOR TO C3B/C4B INACTIVATOR  
 CC (C3BINA) WHICH THEN HYDROLYSES THE COMPLEMENT FRAGMENT C4B. IT  
 CC ALSO ACCELERATES THE DEGRADATION OF THE C4B/C2A COMPLEX (C3  
 CC CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA  
 CC CHAIN BINDS C4B. IT INTERACTS ALSO WITH ANTICOAGULANT PROTEIN S  
 CC AND WITH SERUM AMYLOID P COMPONENT.  
 CC -1- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS  
 CC OF 3 POSSIBLE SORTS: A 570 KDA COMPLEX OF 7 ALPHA CHAINS AND 1  
 CC BETA CHAIN, A 530 KDA HOMOPOLYMER OF ALPHA CHAINS OR A 500 KDA  
 CC COMPLEX OF 6 ALPHA CHAINS AND 1 BETA CHAIN. THE CENTRAL BODY OF  
 CC THE ALPHA CHAIN HOMOPOLYMER SUPPORTS TENTACLES, EACH WITH THE  
 CC BINDING SITE FOR C4B AT THE END.  
 CC -1- TISSUE SPECIFICITY: CHYLOMICRONS IN THE PLASMA.  
 CC -1- SIMILARITY: CONTAINS 8 SUSHI (SCR) REPEATS.  
 CC -1- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.  
 CC -1- CAUTION: IT IS UNCERTAIN WHETHER MER-1 OR MER-17 IS THE INITIATOR.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

identity of PRP as C4b-binding protein.";  
Biochem. Biophys. Res. Commun. 165:138-144(1989).

[2]  
SEQUENCE FROM N.A.  
MEDLINE. 91113199

Aso T., Okamura S., Matsuguchi T., Sakamoto N., Sata T., Niho Y.: Genomic organization of the alpha chain of the human C4b-binding protein. *J. Biol. Chem.* 263: 1011-1015, 1988.

protein gene.";  
Biochem. Biophys. Res. Commun. 174:222-227 (1991).

SEQUENCE OF 9-81 FROM N.A.  
MEDLINE: 88242821  
151

Lintin S.J., Lewin A.R., Reid K.B.M.;  
"Derivation of the sequence of the signal peptide in human

C4b-binding protein and interspecies cross-hybridisation of the C4bp CDNA sequence.";

SEQUENCE OF 203-288 FROM N.A.  
[4]

MEDLINE; 86301119.  
Lintin S.J., Reid K.B.M.;

Studies on the structure of the human C4D-binding protein gene. *J. FEBS Lett.* 204:77-81(1986).

SEQUENCE OF 80-597 FROM N.A.  
MEDLINE; 86025405.

Chung L.P., Bentley D.R., Reid K.B.M.;  
"Molecular cloning and characterization of the cDNA coding for Cab-

biochem. J. 230:133-141(1985).

[6] SEQUENCE OF 49-88.

Chung L.P., Gagnon J., Reid K.B.M.;  
MEDLINE; 85296001.

sequence analysis and alignment of the fragments produced by limited proteolysis with chymotrypsin and the peptides produced by cyanogen bromide.

bromide treatment.";  
Mol. Immunol. 22:427-435(1985).

ELECTRON MICROSCOPY, 3-DIMENSIONAL STRUCTURE, AND LIGAND BINDING.  
MEDLINE: 83221615.

Dahlback B., Smith C.A., Mueller-Eberhard H.J.:  
"Visualization of human C4b-binding protein and its complexes with

Proc. Natl. Acad. Sci. U.S.A. 80:3461-3465(1983).

LONG-TERM CONCENTRATIONS AND CHEMICAL ESTIMATION OF COINTEGRATING  
ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR  
(C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT

ALSO ACCELERATES THE DEGRADATION OF THE C4BC2A COMPLEX (C3 CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA

1. SUBUNIT. DISTUFIDE-LINKED COMPLEX OF CARP ALPHA AND BETA CHAINS WITH SERUM AMYLOID P COMPONENT. AND WITH ANTI COAGULANT PROTEIN S CHAIN BNDS CAR. IT INTERACTS ALSO WITH ANTI COAGULANT PROTEIN S

OF 3 POSSIBLE SORTS: A 570 KDA COMPLEX OF 7 ALPHA CHAINS AND 1 BETA CHAIN, A 530 KDA HOMOHEPTAMER OF ALPHA CHAINS OR A 500 KDA

COMPLEX OF 6 ALPHA CHAINS AND 1 BETA CHAIN. THE CENTRAL BODY OF THE ALPHA CHAIN HOMOPOLYMER SUPPORTS TENTACLES, EACH WITH THE ENDING  $\text{CH}_2\text{OH}$  OR  $\text{CHO}$  END.

- BINDING SITE FOR C4B AT THE END.
- TISSUE SPECIFICITY: CHYLOMICRONS IN THE PLASMA.
- SIMILARITY: CONTAINS 8 SUSHI (SCR) REPEATS.

-1- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.  
-1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-17 IS THE INITIATOR

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the University of Geneva.

the European Bioinformatics Institute. There are no restrictions on use by non-profit institutions as long as its content is in no

modified and this statement is not removed. Usage by and for commentaries requires a license agreement (See <http://www.isb-sib.ch/announcements>)

or send an email to [license@isd-sid.cn](mailto:license@isd-sid.cn)).

```

CC -----
DR EMBL; M31452; AAA36507.1; -
DR EMBL; M62486; AAA36506.1; -
DR EMBL; M62475; AAA36506.1; JOINED.
DR EMBL; M62476; AAA36506.1; JOINED.
DR EMBL; M62477; AAA36506.1; JOINED.
DR EMBL; M62478; AAA36506.1; JOINED.
DR EMBL; M62479; AAA36506.1; JOINED.
DR EMBL; M62480; AAA36506.1; JOINED.
DR EMBL; M62481; AAA36506.1; JOINED.
DR EMBL; M62482; AAA36506.1; JOINED.
DR EMBL; M62484; AAA36506.1; JOINED.
DR EMBL; M62485; AAA36506.1; JOINED.
DR EMBL; M62486; AAA36506.1; JOINED.
DR EMBL; X04284; CAB31244.1; -
DR EMBL; X04296; CA27839.1; -
DR EMBL; X02865; CA26617.1; -
DR PIR; A33568; NBHUC4.
DR HSSP; P10998; LYVC.
DR MIM; 120830; -
DR INTERPRO; IPR000436; -
DR PFAM; PF00084; SUSH1; 8.
KM Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal;
KM Polymorphism.
FT SIGNAL 1 48
FT CHAIN 49 597 C4B-BINDING PROTEIN ALPHA CHAIN.
FT DOMAIN 49 539 8 X SUSHI (SCR) REPEATS.
FT REPEAT 49 109 SUSHI 1.
FT REPEAT 112 171 SUSHI 2.
FT REPEAT 174 235 SUSHI 3.
FT REPEAT 238 295 SUSHI 4.
FT REPEAT 298 361 SUSHI 5.
FT REPEAT 364 423 SUSHI 6.
FT REPEAT 425 481 SUSHI 7.
FT REPEAT 483 539 SUSHI 8.
FT DISULFID 50 96 BY SIMILARITY.
FT DISULFID 81 108 BY SIMILARITY.
FT DISULFID 113 154 BY SIMILARITY.
FT DISULFID 140 170 BY SIMILARITY.
FT DISULFID 175 217 BY SIMILARITY.
FT DISULFID 203 234 BY SIMILARITY.
FT DISULFID 239 281 BY SIMILARITY.
FT DISULFID 297 294 BY SIMILARITY.
FT DISULFID 348 360 BY SIMILARITY.
FT DISULFID 332 360 BY SIMILARITY.
FT DISULFID 7364 7387 BY SIMILARITY.
FT DISULFID 7365 7409 BY SIMILARITY.
FT DISULFID 7399 422 BY SIMILARITY.
FT DISULFID 426 468 BY SIMILARITY.
FT DISULFID 454 480 BY SIMILARITY.
FT DISULFID 484 525 BY SIMILARITY.
FT DISULFID 511 538 BY SIMILARITY.
FT DISULFID 546 546 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT DISULFID 558 558 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT CARBOHYD 221 221 N-LINKED (GLCNAC. . .).
FT CARBOHYD 506 506 N-LINKED (GLCNAC. . .).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .).
FT VARIANT 92
FT VARIANT 357
FT SEQUENCE 597 AA; 67033 MM; 67E03FEA85A16DD CRC64;

```

```

Query Match 19.6%; Score 481.5; DB 1; Length 597;
Best Local Similarity 26.5%; Pred. No. 1.7e-29;
Matches 134; Conservative 78; Mismatches 195; Indels 99; Gaps 25;

```

```

QY 2 DCKGPPPNSEILSGNSEQLYSGTQATYCRPGY-RTLTIVKVC-KNGEWPSNDS 59
DB 49 NCGPPLTSLFAPMDITLTETFRKGTGTLTKYCLPGYVSHSTQTLTCSNDEWYNT-- 106
QY 60 RICRRRPGHEDPTFGSRLAVGSEFEGCAKVYVTCDEGYLLGEIDYRECDAD----G 115

```

```

DB 107 -FCIYKRRHRELNG--QVEIKTDLISFGSIEESCSGFFLLISTSR-CEVDDRGV 162
QY 116 WTNDIPICEVVKCLPVTLENGRIYSGAEPDOEYFGQVREECNSGKLEGQEMHCS 175
DB 163 WSHPLPOCEIYVCKRPDRIKNGR---HSGEENFYAGSHVYSCDPRSLIGHASISCT 218
QY 176 -EN--GLMSNRKPOCVLEISCLPRVENDGTY-LKPYKEMERQYCKQGFYKKEKD 230
DB 219 VENETIGWRPSPPTCEKTKRPPVSHGEMVSGEPILYNKDTIVFCQKGFVL--KCS 276
QY 231 AV--C-TSGWNPQ-PSCSEMTCLF-PYIPNGIY--PHRIK---HRIDETRYCKNGF 280
DB 277 SVIHCDAASKNPPSPACEPNCINLPDIIPASMETYPRPTEDYVVGTVLRYCHRGY 336
QY 281 YPATRSPVSKTYTG--WIPAPRCSLKPDPPQFQKGRLYEESRRP----- 325
DB 337 KPTDEPTTVICQKMLRMTPYGCEALCPPEKLNNGEITQHRKSRPANHCYFYGDEIS 396
QY 326 -----YFVPVIG-----KEYSYCDN 341
DB 397 FSCHESTRFSAICQDGTWSPPTPCGDCINPPKIAHGHYKSSSYSPFEKEIIECDK 456
QY 342 GPTTSPQSYWDYLRCTVNGMEDEVP-CLRQCIFHYVEGESSYWMORRYEGQSAKVQCHS 400
DB 457 GYIIVGQA---KLSCSYMSAPAPQCKALCKRPVELNGLRSLVDQYEPENVTIQCDS 513
QY 401 GYSLPNGDPTTYCTEN-GWSP-PRC 424
DB 514 GYGAV-GPQSTICSGNRMTYPEVPC 538

RESULT 4
C4BP_BOVIN STANDARD; PRT; 610 AA.
ID C4BP_BOVIN 028065;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE C4B-BINDING PROTEIN ALPHA CHAIN PRECURSOR.
GN C4BPA.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Bovinae; Bos.
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE: 95015909.
RA Hillarp A., Thern A., Dahlbaeck B.;
RT "Bovine C4b binding protein. Molecular cloning of the alpha- and
RT beta-chains provides structural background for lack of complex
RT formation with protein S."
RL J. Immunol. 153:4190-4199(1994).
CC -!- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
CC ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR
CC (C3BIN), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT
CC ALSO ACCELERATES THE DEGRADATION OF THE C4B2A COMPLEX (C3
CC CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA
CC CHAIN BINDS C4B. IT INTERACTS ALSO WITH SERUM AMYLOID P COMPONENT.
CC -!- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS
CC (BY SIMILARITY)
CC -!- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.
CC -!- SIMILARITY: CONTAINS 8 SUSHI (SCR) REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```



Query Match	Best Local Similarity	Matches	137; Conservative	Score 449; DB 1; Length 558; Pred. No. 4,6e-27; Mismatches 198; Indels 108; Gaps 24
FT DISULEFID	2328	2351	BY SIMILARITY.	
FT DISULEFID	2329	2373	BY SIMILARITY.	
FT DISULEFID	2363	386	BY SIMILARITY.	
FT DISULEFID	390	431	BY SIMILARITY.	
FT DISULEFID	417	443	BY SIMILARITY.	
FT DISULEFID	447	488	BY SIMILARITY.	
FT DISULEFID	474	501	BY SIMILARITY.	
FT DISULEFID	509	509	INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).	
FT DISULEFID	521	521	INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).	
FT CAROHND	31	31	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT CAROHND	177	177	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT CAROHND	186	186	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT CAROHND	459	469	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT CAROHND	491	491	N-LINKED (GLCNAC. . .) (POTENTIAL).	
SO SEQUENCE	558 AA;	62266 MW;	592F0C667ED1E5FF CRC64;	
Query Match				
Best Local Similarity 26.7%; Pred. No. 4,6e-27;				
Matches 137; Conservative 71; Mismatches 198; Indels 108; Gaps 24				
QY 5	GPPRRNSIIILGSMSEQLSECTQATYTCRGQY	RTLGTIVAKCKN	GEWYPSNDRIC	62
DB 16	GPPDLPALPAASENNQDTEESHHTLTRYRCRGYSNASSOSLYCPDLKW	---	QIINIK	72
QY 63	RRPGCHGPDTPFGSFRLVAGSEFEFGAKVAVVTCDEGYLLG	---	EDYRRCADG	---
DB 73	VAKSCRNPDLLNG	---	KVEYKTDPLFEGSIEFSCSGYLLISSHSYCYCIQKGYMSDP	130
QY 120	IPICEVYKCLPYTELENGHIVSGAAEPDQYVFGOVAFRECSGFKIEGQKEMHCS	---	---	175
DB 131	LPCEVIYAKCGMPDINSNG	---	HNGREEFFYRBSVYKCDPDTLLGNASITCTVYVK	187
QY 176	ENGLSMNEKPOCVELISCLPRVYENG	---	GIYIKPVYKKNRPYKKGQGVYKERDAVCT	234
DB 188	TYGVWSPSPPTCERILICPMFKVLHGINSGFHTYTKDSYVFVQCKGVLGSGVHCE	---	---	247
QY 235	GSQ-WNPOPSCEMCL	---	TPYIPNG	---
DB 248	ADGSMSPVYCEINSCDIPIDIPNALITSPRPREDYVPQTVIARICRPYEPATRP	---	---	307
QY 268	VSKICL	---	IGVLPAPRCSLKPCDFPOFKIGRLYIESRRP	---
DB 308	M--TVICQKDLWSMLRCKCEICCPVPDPKSVYIOHKAHPDMDCTYF	---	---	361
QY 338	YCDN	---	---	355
DB 362	TCQNDIMLTATCKSDGTWHPRTPSCHOSCDFPAPALAHGYTKSSSY	---	YVRQVYEEBE	419
QY 356	---	---	CTVNGMEVEVP	---
DB 420	GYYLVGEATISQWYQMPAPAPQCKALCRKPEIGNGVLTNDQYVEVETNI	---	IQCDSGV	479
QY 404	LPNGQDYVYCTENG	---	WSP	---
DB 480	M-LGSOSITCSENGTYWYFVRSKCEQDEYPRDCEHV	---	---	512
RESULT 7				
ID	CRL_HUMAN	STANDARD:	PRT:	2039 AA.
AC	P17527,			
DT	01-NOV-1990 (Rel. 16, Created)			
DT	01-NOV-1990 (Rel. 16, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	COMPLEMENT RECEPTOR TYPE 1 PRECURSOR (C3B/4AB RECEPTOR) (CD35			
DE	ANTIGEN).			
GN	CRI OR C3BR.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
NP	(1)			
NP	SEQUENCE FROM N.A.			

```

RX MEDLINE: 89035992.
RA Klinkstein L.B., Bartow T.J., Miletic V., Rabson L.D., Smith J.A.,
RA Fearon D.T.;
RT "Identification of distinct C3b and C4b recognition sites in the
RT human C3b/C4b receptor (CRI, CD35) by deletion mutagenesis.";
RL J. Exp. Med. 168:1699-1717(1988).
RN [2]
RP SEQUENCE OF 503-2039 FROM N.A.
RX MEDLINE: 87168191.
RA Klinkstein L.B., Wong W.W., Smith J.A., Weis J.H., Wilson J.G.,
RA Fearon D.T.;
RT "Human C3b/C4b receptor (CRI). Demonstration of long homologous
RT repeating domains that are composed of the short consensus repeats
RT characteristics of C3/C4 binding proteins.";
RL J. Exp. Med. 165:1095-1112(1987).
RN [3]
RP SEQUENCE OF 761-783; 831-845 AND 1179-1195 FROM N.A.
RX MEDLINE: 86067975.
RA Wong W.W., Klinkstein L.B., Smith J.A., Weis J.H., Fearon D.T.;
RT "Identification of a partial cDNA clone for the human receptor for
RT Complement fragments C3b/C4b.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:7711-7715(1985).
CC -I- FUNCTION: CRI, PRESENT ON ERYTHROCYTES, LEUCOCYTES, GLOMERULAR
CC PODOCYTES, AND SPLEEN FOLLICULAR DENDRITIC CELLS, MEDIATES THE
CC BINDING BY THESE CELLS OF PARTICLES AND IMMUNE COMPLEXES THAT HAVE
CC ACTIVATED COMPLEMENT.
CC -I- SUBUNIT: MONOMER.
CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -I- POLYMORPHISM: CRI IS RESPONSIBLE FOR THE KNOBS BLOOD GROUP SYSTEM.
CC -I- MISCELLANEOUS: SEVEN SHORT CONSENSUS REPEATS (SCR) CONSTITUTE A
CC LONG HOMOLOGOUS REPEAT (LHR). THE N-TERMINAL TWO SCRS OF LHR-A
CC CONTAINED A SITE DETERMINING C4 SPECIFICITY, AND THE N-TERMINAL
CC TWO SCRS OF LHR-B AND -C EACH HAD A SITE DETERMINING C3
CC SPECIFICITY.
CC -I- MISCELLANEOUS: THIS IS THE SEQUENCE OF THE F ALLOTYPE OF CRI.
CC -I- SIMILARITY: CONTAINS 30 SUSHI (SCR) REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M11569; AAA52297.1; -
DR EMBL: M11617; AAA52298.1; -
DR EMBL: M11618; AAA52299.1; -
DR EMBL: Y00816; CAA68755.1; -
DR EMBL: X05309; CAA28933.1; -
DR PIR: A28507; A28507.
DR PIR: A24748; A24748.
DR PIR: B24748; B24748.
DR PIR: C24748; C24748.
DR PIR: S03843; S03843.
DR HSSP: P08603; IHFI.
DR MIM: 120620; -.
DR INTERPRO: IPR000436; -.
DR PRAM: PR000084; Sush1; 30.
KW Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
KW Receptor; Sush1; Blood group antigen.
FT SIGNAL 1 41
FT CHAIN 1 41
FT DOMAIN 42 2039
FT TRANSMEM 1972 1996
FT DOMAIN 1997 2039
FT MOD_RES 42 42
FT DOMAIN 42 488
FT REPEAT 42 100
FT REPEAT 103 162
FT REPEAT 165 233
FT REPEAT 237 294
FT SUSHI A1.
FT SUSHI A2.
FT SUSHI A3.
FT SUSHI A4.
FT SUSHI A5.
FT SUSHI A6.
FT SUSHI A7.
FT SUSHI A8.
FT SUSHI A9.
FT SUSHI A10.
FT SUSHI A11.
FT SUSHI A12.
FT SUSHI A13.
FT SUSHI A14.
FT SUSHI A15.
FT SUSHI A16.
FT SUSHI A17.
FT SUSHI A18.
FT SUSHI A19.
FT SUSHI A20.
FT SUSHI A21.
FT SUSHI A22.
FT SUSHI A23.
FT SUSHI A24.
FT SUSHI A25.
FT SUSHI A26.
FT SUSHI A27.
FT SUSHI A28.
FT SUSHI A29.
FT SUSHI A30.
FT SUSHI A31.
FT SUSHI A32.
FT SUSHI A33.
FT SUSHI A34.
FT SUSHI A35.
FT SUSHI A36.
FT SUSHI A37.
FT SUSHI A38.
FT SUSHI A39.
FT SUSHI A40.
FT SUSHI A41.
FT SUSHI A42.
FT SUSHI A43.
FT SUSHI A44.
FT SUSHI A45.
FT SUSHI A46.
FT SUSHI A47.
FT SUSHI A48.
FT SUSHI A49.
FT SUSHI A50.
FT SUSHI A51.
FT SUSHI A52.
FT SUSHI A53.
FT SUSHI A54.
FT SUSHI A55.
FT SUSHI A56.
FT SUSHI A57.
FT SUSHI A58.
FT SUSHI A59.
FT SUSHI A60.
FT SUSHI A61.
FT SUSHI A62.
FT SUSHI A63.
FT SUSHI A64.
FT SUSHI A65.
FT SUSHI A66.
FT SUSHI A67.
FT SUSHI A68.
FT SUSHI A69.
FT SUSHI A70.
FT SUSHI A71.
FT SUSHI A72.
FT SUSHI A73.
FT SUSHI A74.
FT SUSHI A75.
FT SUSHI A76.
FT SUSHI A77.
FT SUSHI A78.
FT SUSHI A79.
FT SUSHI A80.
FT SUSHI A81.
FT SUSHI A82.
FT SUSHI A83.
FT SUSHI A84.
FT SUSHI A85.
FT SUSHI A86.
FT SUSHI A87.
FT SUSHI A88.
FT SUSHI A89.
FT SUSHI A90.
FT SUSHI A91.
FT SUSHI A92.
FT SUSHI A93.
FT SUSHI A94.
FT SUSHI A95.
FT SUSHI A96.
FT SUSHI A97.
FT SUSHI A98.
FT SUSHI A99.
FT SUSHI A100.
FT SUSHI A101.
FT SUSHI A102.
FT SUSHI A103.
FT SUSHI A104.
FT SUSHI A105.
FT SUSHI A106.
FT SUSHI A107.
FT SUSHI A108.
FT SUSHI A109.
FT SUSHI A110.
FT SUSHI A111.
FT SUSHI A112.
FT SUSHI A113.
FT SUSHI A114.
FT SUSHI A115.
FT SUSHI A116.
FT SUSHI A117.
FT SUSHI A118.
FT SUSHI A119.
FT SUSHI A120.
FT SUSHI A121.
FT SUSHI A122.
FT SUSHI A123.
FT SUSHI A124.
FT SUSHI A125.
FT SUSHI A126.
FT SUSHI A127.
FT SUSHI A128.
FT SUSHI A129.
FT SUSHI A130.
FT SUSHI A131.
FT SUSHI A132.
FT SUSHI A133.
FT SUSHI A134.
FT SUSHI A135.
FT SUSHI A136.
FT SUSHI A137.
FT SUSHI A138.
FT SUSHI A139.
FT SUSHI A140.
FT SUSHI A141.
FT SUSHI A142.
FT SUSHI A143.
FT SUSHI A144.
FT SUSHI A145.
FT SUSHI A146.
FT SUSHI A147.
FT SUSHI A148.
FT SUSHI A149.
FT SUSHI A150.
FT SUSHI A151.
FT SUSHI A152.
FT SUSHI A153.
FT SUSHI A154.
FT SUSHI A155.
FT SUSHI A156.
FT SUSHI A157.
FT SUSHI A158.
FT SUSHI A159.
FT SUSHI A160.
FT SUSHI A161.
FT SUSHI A162.
FT SUSHI A163.
FT SUSHI A164.
FT SUSHI A165.
FT SUSHI A166.
FT SUSHI A167.
FT SUSHI A168.
FT SUSHI A169.
FT SUSHI A170.
FT SUSHI A171.
FT SUSHI A172.
FT SUSHI A173.
FT SUSHI A174.
FT SUSHI A175.
FT SUSHI A176.
FT SUSHI A177.
FT SUSHI A178.
FT SUSHI A179.
FT SUSHI A180.
FT SUSHI A181.
FT SUSHI A182.
FT SUSHI A183.
FT SUSHI A184.
FT SUSHI A185.
FT SUSHI A186.
FT SUSHI A187.
FT SUSHI A188.
FT SUSHI A189.
FT SUSHI A190.
FT SUSHI A191.
FT SUSHI A192.
FT SUSHI A193.
FT SUSHI A194.
FT SUSHI A195.
FT SUSHI A196.
FT SUSHI A197.
FT SUSHI A198.
FT SUSHI A199.
FT SUSHI A200.
FT SUSHI A201.
FT SUSHI A202.
FT SUSHI A203.
FT SUSHI A204.
FT SUSHI A205.
FT SUSHI A206.
FT SUSHI A207.
FT SUSHI A208.
FT SUSHI A209.
FT SUSHI A210.
FT SUSHI A211.
FT SUSHI A212.
FT SUSHI A213.
FT SUSHI A214.
FT SUSHI A215.
FT SUSHI A216.
FT SUSHI A217.
FT SUSHI A218.
FT SUSHI A219.
FT SUSHI A220.
FT SUSHI A221.
FT SUSHI A222.
FT SUSHI A223.
FT SUSHI A224.
FT SUSHI A225.
FT SUSHI A226.
FT SUSHI A227.
FT SUSHI A228.
FT SUSHI A229.
FT SUSHI A230.
FT SUSHI A231.
FT SUSHI A232.
FT SUSHI A233.
FT SUSHI A234.
FT SUSHI A235.
FT SUSHI A236.
FT SUSHI A237.
FT SUSHI A238.
FT SUSHI A239.
FT SUSHI A240.
FT SUSHI A241.
FT SUSHI A242.
FT SUSHI A243.
FT SUSHI A244.
FT SUSHI A245.
FT SUSHI A246.
FT SUSHI A247.
FT SUSHI A248.
FT SUSHI A249.
FT SUSHI A250.
FT SUSHI A251.
FT SUSHI A252.
FT SUSHI A253.
FT SUSHI A254.
FT SUSHI A255.
FT SUSHI A256.
FT SUSHI A257.
FT SUSHI A258.
FT SUSHI A259.
FT SUSHI A260.
FT SUSHI A261.
FT SUSHI A262.
FT SUSHI A263.
FT SUSHI A264.
FT SUSHI A265.
FT SUSHI A266.
FT SUSHI A267.
FT SUSHI A268.
FT SUSHI A269.
FT SUSHI A270.
FT SUSHI A271.
FT SUSHI A272.
FT SUSHI A273.
FT SUSHI A274.
FT SUSHI A275.
FT SUSHI A276.
FT SUSHI A277.
FT SUSHI A278.
FT SUSHI A279.
FT SUSHI A280.
FT SUSHI A281.
FT SUSHI A282.
FT SUSHI A283.
FT SUSHI A284.
FT SUSHI A285.
FT SUSHI A286.
FT SUSHI A287.
FT SUSHI A288.
FT SUSHI A289.
FT SUSHI A290.
FT SUSHI A291.
FT SUSHI A292.
FT SUSHI A293.
FT SUSHI A294.
FT SUSHI A295.
FT SUSHI A296.
FT SUSHI A297.
FT SUSHI A298.
FT SUSHI A299.
FT SUSHI A300.
FT SUSHI A301.
FT SUSHI A302.
FT SUSHI A303.
FT SUSHI A304.
FT SUSHI A305.
FT SUSHI A306.
FT SUSHI A307.
FT SUSHI A308.
FT SUSHI A309.
FT SUSHI A310.
FT SUSHI A311.
FT SUSHI A312.
FT SUSHI A313.
FT SUSHI A314.
FT SUSHI A315.
FT SUSHI A316.
FT SUSHI A317.
FT SUSHI A318.
FT SUSHI A319.
FT SUSHI A320.
FT SUSHI A321.
FT SUSHI A322.
FT SUSHI A323.
FT SUSHI A324.
FT SUSHI A325.
FT SUSHI A326.
FT SUSHI A327.
FT SUSHI A328.
FT SUSHI A329.
FT SUSHI A330.
FT SUSHI A331.
FT SUSHI A332.
FT SUSHI A333.
FT SUSHI A334.
FT SUSHI A335.
FT SUSHI A336.
FT SUSHI A337.
FT SUSHI A338.
FT SUSHI A339.
FT SUSHI A340.
FT SUSHI A341.
FT SUSHI A342.
FT SUSHI A343.
FT SUSHI A344.
FT SUSHI A345.
FT SUSHI A346.
FT SUSHI A347.
FT SUSHI A348.
FT SUSHI A349.
FT SUSHI A350.
FT SUSHI A351.
FT SUSHI A352.
FT SUSHI A353.
FT SUSHI A354.
FT SUSHI A355.
FT SUSHI A356.
FT SUSHI A357.
FT SUSHI A358.
FT SUSHI A359.
FT SUSHI A360.
FT SUSHI A361.
FT SUSHI A362.
FT SUSHI A363.
FT SUSHI A364.
FT SUSHI A365.
FT SUSHI A366.
FT SUSHI A367.
FT SUSHI A368.
FT SUSHI A369.
FT SUSHI A370.
FT SUSHI A371.
FT SUSHI A372.
FT SUSHI A373.
FT SUSHI A374.
FT SUSHI A375.
FT SUSHI A376.
FT SUSHI A377.
FT SUSHI A378.
FT SUSHI A379.
FT SUSHI A380.
FT SUSHI A381.
FT SUSHI A382.
FT SUSHI A383.
FT SUSHI A384.
FT SUSHI A385.
FT SUSHI A386.
FT SUSHI A387.
FT SUSHI A388.
FT SUSHI A389.
FT SUSHI A390.
FT SUSHI A391.
FT SUSHI A392.
FT SUSHI A393.
FT SUSHI A394.
FT SUSHI A395.
FT SUSHI A396.
FT SUSHI A397.
FT SUSHI A398.
FT SUSHI A399.
FT SUSHI A400.
FT SUSHI A401.
FT SUSHI A402.
FT SUSHI A403.
FT SUSHI A404.
FT SUSHI A405.
FT SUSHI A406.
FT SUSHI A407.
FT SUSHI A408.
FT SUSHI A409.
FT SUSHI A410.
FT SUSHI A411.
FT SUSHI A412.
FT SUSHI A413.
FT SUSHI A414.
FT SUSHI A415.
FT SUSHI A416.
FT SUSHI A417.
FT SUSHI A418.
FT SUSHI A419.
FT SUSHI A420.
FT SUSHI A421.
FT SUSHI A422.
FT SUSHI A423.
FT SUSHI A424.
FT SUSHI A425.
FT SUSHI A426.
FT SUSHI A427.
FT SUSHI A428.
FT SUSHI A429.
FT SUSHI A430.
FT SUSHI A431.
FT SUSHI A432.
FT SUSHI A433.
FT SUSHI A434.
FT SUSHI A435.
FT SUSHI A436.
FT SUSHI A437.
FT SUSHI A438.
FT SUSHI A439.
FT SUSHI A440.
FT SUSHI A441.
FT SUSHI A442.
FT SUSHI A443.
FT SUSHI A444.
FT SUSHI A445.
FT SUSHI A446.
FT SUSHI A447.
FT SUSHI A448.
FT SUSHI A449.
FT SUSHI A450.
FT SUSHI A451.
FT SUSHI A452.
FT SUSHI A453.
FT SUSHI A454.
FT SUSHI A455.
FT SUSHI A456.
FT SUSHI A457.
FT SUSHI A458.
FT SUSHI A459.
FT SUSHI A460.
FT SUSHI A461.
FT SUSHI A462.
FT SUSHI A463.
FT SUSHI A464.
FT SUSHI A465.
FT SUSHI A466.
FT SUSHI A467.
FT SUSHI A468.
FT SUSHI A469.
FT SUSHI A470.
FT SUSHI A471.
FT SUSHI A472.
FT SUSHI A473.
FT SUSHI A474.
FT SUSHI A475.
FT SUSHI A476.
FT SUSHI A477.
FT SUSHI A478.
FT SUSHI A479.
FT SUSHI A480.
FT SUSHI A481.
FT SUSHI A482.
FT SUSHI A483.
FT SUSHI A484.
FT SUSHI A485.
FT SUSHI A486.
FT SUSHI A487.
FT SUSHI A488.
FT SUSHI A489.
FT SUSHI A490.
FT SUSHI A491.
FT SUSHI A492.
FT SUSHI A493.
FT SUSHI A494.
FT SUSHI A495.
FT SUSHI A496.
FT SUSHI A497.
FT SUSHI A498.
FT SUSHI A499.
FT SUSHI A500.
FT SUSHI A501.
FT SUSHI A502.
FT SUSHI A503.
FT SUSHI A504.
FT SUSHI A505.
FT SUSHI A506.
FT SUSHI A507.
FT SUSHI A508.
FT SUSHI A509.
FT SUSHI A510.
FT SUSHI A511.
FT SUSHI A512.
FT SUSHI A513.
FT SUSHI A514.
FT SUSHI A515.
FT SUSHI A516.
FT SUSHI A517.
FT SUSHI A518.
FT SUSHI A519.
FT SUSHI A520.
FT SUSHI A521.
FT SUSHI A522.
FT SUSHI A523.
FT SUSHI A524.
FT SUSHI A525.
FT SUSHI A526.
FT SUSHI A527.
FT SUSHI A528.
FT SUSHI A529.
FT SUSHI A530.
FT SUSHI A531.
FT SUSHI A532.
FT SUSHI A533.
FT SUSHI A534.
FT SUSHI A535.
FT SUSHI A536.
FT SUSHI A537.
FT SUSHI A538.
FT SUSHI A539.
FT SUSHI A540.
FT SUSHI A541.
FT SUSHI A542.
FT SUSHI A543.
FT SUSHI A544.
FT SUSHI A545.
FT SUSHI A546.
FT SUSHI A547.
FT SUSHI A548.
FT SUSHI A549.
FT SUSHI A550.
FT SUSHI A551.
FT SUSHI A552.
FT SUSHI A553.
FT SUSHI A554.
FT SUSHI A555.
FT SUSHI A556.
FT SUSHI A557.
FT SUSHI A558.
FT SUSHI A559.
FT SUSHI A560.
FT SUSHI A561.
FT SUSHI A562.
FT SUSHI A563.
FT SUSHI A564.
FT SUSHI A565.
FT SUSHI A566.
FT SUSHI A567.
FT SUSHI A568.
FT SUSHI A569.
FT SUSHI A570.
FT SUSHI A571.
FT SUSHI A572.
FT SUSHI A573.
FT SUSHI A574.
FT SUSHI A575.
FT SUSHI A576.
FT SUSHI A577.
FT SUSHI A578.
FT SUSHI A579.
FT SUSHI A580.
FT SUSHI A581.
FT SUSHI A582.
FT SUSHI A583.
FT SUSHI A584.
FT SUSHI A585.
FT SUSHI A586.
FT SUSHI A587.
FT SUSHI A588.
FT SUSHI A589.
FT SUSHI A590.
FT SUSHI A591.
FT SUSHI A592.
FT SUSHI A593.
FT SUSHI A594.
FT SUSHI A595.
FT SUSHI A596.
FT SUSHI A597.
FT SUSHI A598.
FT SUSHI A599.
FT SUSHI A600.
FT SUSHI A601.
FT SUSHI A602.
FT SUSHI A603.
FT SUSHI A604.
FT SUSHI A605.
FT SUSHI A606.
FT SUSHI A607.
FT SUSHI A608.
FT SUSHI A609.
FT SUSHI A610.
FT SUSHI A611.
FT SUSHI A612.
FT SUSHI A613.
FT SUSHI A614.
FT SUSHI A615.
FT SUSHI A616.
FT SUSHI A617.
FT SUSHI A618.
FT SUSHI A619.
FT SUSHI A620.
FT SUSHI A621.
FT SUSHI A622.
FT SUSHI A623.
FT SUSHI A624.
FT SUSHI A625.
FT SUSHI A626.
FT SUSHI A627.
FT SUSHI A628.
FT SUSHI A629.
FT SUSHI A630.
FT SUSHI A631.
FT SUSHI A632.
FT SUSHI A633.
FT SUSHI A634.
FT SUSHI A635.
FT SUSHI A636.
FT SUSHI A637.
FT SUSHI A638.
FT SUSHI A639.
FT SUSHI A640.
FT SUSHI A641.
FT SUSHI A642.
FT SUSHI A643.
FT SUSHI A644.
FT SUSHI A645.
FT SUSHI A646.
FT SUSHI A647.
FT SUSHI A648.
FT SUSHI A649.
FT SUSHI A650.
FT SUSHI A651.
FT SUSHI A652.
FT SUSHI A653.
FT SUSHI A654.
FT SUSHI A655.
FT SUSHI A656.
FT SUSHI A657.
FT SUSHI A658.
FT SUSHI A659.
FT SUSHI A660.
FT SUSHI A661.
FT SUSHI A662.
FT SUSHI A663.
FT SUSHI A664.
FT SUSHI A665.
FT SUSHI A666.
FT SUSHI A667.
FT SUSHI A668.
FT SUSHI A669.
FT SUSHI A670.
FT SUSHI A671.
FT SUSHI A672.
FT SUSHI A673.
FT SUSHI A674.
FT SUSHI A675.
FT SUSHI A676.
FT SUSHI A677.
FT SUSHI A678.
FT SUSHI A679.
FT SUSHI A680.
FT SUSHI A681.
FT SUSHI A682.
FT SUSHI A683.
FT SUSHI A684.
FT SUSHI A685.
FT SUSHI A686.
FT SUSHI A687.
FT SUSHI A688.
FT SUSHI A689.
FT SUSHI A690.
FT SUSHI A691.
FT SUSHI A692.
FT SUSHI A693.
FT SUSHI A694.
FT SUSHI A695.
FT SUSHI A696.
FT SUSHI A697.
FT SUSHI A698.
FT SUSHI A699.
FT SUSHI A700.
FT SUSHI A701.
FT SUSHI A702.
FT SUSHI A703.
FT SUSHI A704.
FT SUSHI A705.
FT SUSHI A706.
FT SUSHI A707.
FT SUSHI A708.
FT SUSHI A709.
FT SUSHI A710.
FT SUSHI A711.
FT SUSHI A712.
FT SUSHI A713.
FT SUSHI A714.
FT SUSHI A715.
FT SUSHI A716.
FT SUSHI A717.
FT SUSHI A718.
FT SUSHI A719.
FT SUSHI A720.
FT SUSHI A721.
FT SUSHI A722.
FT SUSHI A723.
FT SUSHI A724.
FT SUSHI A725.
FT SUSHI A726.
FT SUSHI A727.
FT SUSHI A728.
FT SUSHI A729.
FT SUSHI A730.
FT SUSHI A731.
FT SUSHI A732.
FT SUSHI A733.
FT SUSHI A734.
FT SUSHI A735.
FT SUSHI A736.
FT SUSHI A737.
FT SUSHI A738.
FT SUSHI A739.
FT SUSHI A740.
FT SUSHI A741.
FT SUSHI A742.
FT SUSHI A743.
FT SUSHI A744.
FT SUSHI A745.
FT SUSHI A746.
FT SUSHI A747.
FT SUSHI A748.
FT SUSHI A749.
FT SUSHI A750.
FT SUSHI A751.
FT SUSHI A752.
FT SUSHI A753.
FT SUSHI A754.
FT SUSHI A755.
FT SUSHI A756.
FT SUSHI A757.
FT SUSHI A758.
FT SUSHI A759.
FT SUSHI A760.
FT SUSHI A761.
FT SUSHI A762.
FT SUSHI A763.
FT SUSHI A764.
FT SUSHI A765.
FT SUSHI A766.
FT SUSHI A767.
FT SUSHI A768.
FT SUSHI A769.
FT SUSHI A770.
FT SUSHI A771.
FT SUSHI A772.
FT SUSHI A773.
FT SUSHI A774.
FT SUSHI A775.
FT SUSHI A776.
FT SUSHI A777.
FT SUSHI A778.
FT SUSHI A779.
FT SUSHI A780.
FT SUSHI A781.
FT SUSHI A782.
FT SUSHI A783.
FT SUSHI A784.
FT SUSHI A785.
FT SUSHI A786.
FT SUSHI A787.
FT SUSHI A788.
FT SUSHI A789.
FT SUSHI A790.
FT SUSHI A791.
FT SUSHI A792.
FT SUSHI A793.
FT SUSHI A794.
FT SUSHI A795.
FT SUSHI A796.
FT SUSHI A797.
FT SUSHI A798.
FT SUSHI A799.
FT SUSHI A800.
FT SUSHI A801.
FT SUSHI A802.
FT SUSHI A803.
FT SUSHI A804.
FT SUSHI A805.
FT SUSHI A806.
FT SUSHI A807.
FT SUSHI A808.
FT SUSHI A809.
FT SUSHI A810.
FT SUSHI A811.
FT SUSHI A812.
FT SUSHI A813.
FT SUSHI A814.
FT SUSHI A815.
FT SUSHI A816.
FT SUSHI A817.
FT SUSHI A818.
FT SUSHI A819.
FT SUSHI A820.
FT SUSHI A821.
FT SUSHI A822.
FT SUSHI A823.
FT SUSHI A824.
FT SUSHI A825.
FT SUSHI A826.
FT SUSHI A827.
FT SUSHI A828.
FT SUSHI A829.
FT SUSHI A830.
FT SUSHI A831.
FT SUSHI A832.
FT SUSHI A833.
FT SUSHI A834.
FT SUSHI A835.
FT SUSHI A836.
FT SUSHI A837.
FT SUSHI A838.
FT SUSHI A839.
FT SUSHI A840.
FT SUSHI A841.
FT SUSHI A842.
FT SUSHI A843.
FT SUSHI A844.
FT SUSHI A845.
FT SUSHI A846.
FT SUSHI A847.
FT SUSHI A848.
FT SUSHI A849.
FT SUSHI A850.
FT SUSHI A851.
FT SUSHI A852.
FT SUSHI A853.
FT SUSHI A854.
FT SUSHI A855.
FT SUSHI A856.
FT SUSHI A857.
FT SUSHI A858.
FT SUSHI A859.
FT SUSHI A860.
FT SUSHI A861.
FT SUSHI A862.
FT SUSHI A863.
FT SUSHI A864.
FT SUSHI A865.
FT SUSHI A866.
FT SUSHI A867.
FT SUSHI A868.
FT SUSHI A869.
FT SUSHI A870.
FT SUSHI A871.
FT SUSHI A872.
FT SUSHI A873.
FT SUSHI A874.
FT SUSHI A875.
FT SUSHI A876.
FT SUSHI A877.
FT SUSHI A878.
FT SUSHI A879.
FT SUSHI A880.
FT SUSHI A881.
FT SUSHI A882.
FT SUSHI A883.
FT SUSHI A884.
FT SUSHI A885.
FT SUSHI A886.
FT SUSHI A887.
FT SUSHI A888.
FT SUSHI A889.
FT SUSHI A890.
FT SUSHI A891.
FT SUSHI A892.
FT SUSHI A893.
FT SUSHI A894.
FT SUSHI A895.
FT SUSHI A896.
FT SUSHI A897.
FT SUSHI A898.
FT SUSHI A899.
FT SUSHI A900.
FT SUSHI A901.
FT SUSHI A902.
FT SUSHI A903.
FT SUSHI A904.
FT SUSHI A905.
FT SUSHI A906.
FT SUSHI A907.
FT SUSHI A908.
FT SUSHI A909.
FT SUSHI A910.
FT SUSHI A911.
FT SUSHI A912.
FT SUSHI A913.
FT SUSHI A914.
FT SUSHI A915.
FT SUSHI A916.
FT SUSHI A917.
FT SUSHI A918.
FT SUSHI A919.
FT SUSHI A920.
FT SUSHI A921.
FT SUSHI A922.
FT SUSHI A923.
FT SUSHI A924.
FT SUSHI A925.
FT SUSHI A926.
FT SUSHI A927.
FT SUSHI A928.
FT SUSHI A929.
FT SUSHI A930.
FT SUSHI A931.
FT SUSHI A932.
FT SUSHI A933.
FT SUSHI A934.
FT SUSHI A935.
FT SUSHI A936.
FT SUSHI A937.
FT SUSHI A938.
FT SUSHI A939.
FT SUSHI A940.
FT SUSHI A941.
FT SUSHI A942.
FT SUSHI A943.
FT SUSHI A944.
FT SUSHI A945.
FT SUSHI A946.
FT SUSHI A947.
FT SUSHI A948.
FT SUSHI A949.
FT SUSHI A950.
FT SUSHI A951.
FT SUSHI A952.
FT SUSHI A953.
FT SUSHI A954.
FT SUSHI A955.
FT SUSHI A956.
FT SUSHI A957.
FT SUSHI A958.
FT SUSHI A959.
FT SUSHI A960.
FT SUSHI A961.
FT SUSHI A962.
FT SUSHI A963.
FT SUSHI A964.
FT SUSHI A965.
FT SUSHI A966.
FT SUSHI A967.
FT SUSHI A968.
FT SUSHI A969.
FT SUSHI A970.
FT SUSHI A971.
FT SUSHI A972.
FT SUSHI A973.
FT SUSHI A974.
FT SUSHI A975.
FT SUSHI A976.
FT SUSHI A977.
FT SUSHI A978.
FT SUSHI A979.
FT SUSHI A980.
FT SUSHI A981.
FT SUSHI A982.
FT SUSHI A983.
FT SUSHI A984.
FT SUSHI A985.
FT SUSHI A986.
FT SUSHI A987.
FT SUSHI A988.
FT SUSHI A989.
FT SUSHI A990.
FT SUSHI A991.
FT SUSHI A992.
FT SUSHI A993.
FT SUSHI A994.
FT SUSHI A995.
FT SUSHI A996.
FT SUSHI A997.
FT SUSHI A998.
FT SUSHI A999.
FT SUSHI A1000.
FT SUSHI A1001.
FT SUSHI A1002.
FT SUSHI A1003.
FT SUSHI A1004.
FT SUSHI A1005.
FT SUSHI A1006.
FT SUSHI A1007.
FT SUSHI A1008.
FT SUSHI A1009.
FT SUSHI A1010.
FT SUSHI A1011.
FT SUSHI A1012.
FT SUSHI A1013.
FT SUSHI A1014.
FT SUSHI A1015.
FT SUSHI A1016.
FT SUSHI A1017.
FT SUSHI A1018.
FT SUSHI A1019.
FT SUSHI A1020.
FT SUSHI A1021.
FT SUSHI A1022.
FT SUSHI A1023.
FT SUSHI A1024.
FT SUSHI A1025.
FT SUSHI A1026.
FT SUSHI A1027.
FT SUSHI A1028.
FT SUSHI A1029.
FT SUSHI A1030.
FT SUSHI A1031.
FT SUSHI A1032.
FT SUSHI A1033.
FT SUSHI A1034.
FT SUSHI A1035.
FT SUSHI A1036.
FT SUSHI A1037.
FT SUSHI A1038.
FT SUSHI A1039.
FT SUSHI A1040.
FT SUSHI A1041.
FT SUSHI A1042.
FT SUSHI A1043.
FT SUSHI A1044.
FT SUSHI A1045.
FT SUSHI A1046.
FT SUSHI A1047.
FT SUSHI A1048.
FT SUSHI A1049.
FT SUSHI A1050.
FT SUSHI A1051.
FT SUSHI A1052.
FT SUSHI A1053.
FT SUSHI A1054.
FT SUSHI A1055.
FT SUSHI A1056.
FT SUSHI A1057.
FT SUSHI A1058.
FT SUSHI A1059.
FT SUSHI A1060.
FT SUSHI A1061.
FT SUSHI A1062.
FT SUSHI A1063.
FT SUSHI A1064.
FT SUSHI A1065.
FT SUSHI A1066.
FT SUSHI A1067.
FT SUSHI A1068.
FT SUSHI A1069.
FT SUSHI A1070.
FT SUSHI A1071.
FT SUSHI A1072.
FT SUSHI A1073.
FT SUSHI A1074.
FT SUSHI A1075.
FT SUSHI A1076.
FT SUSHI A1077.
FT SUSHI A1078.
FT SUSHI A1079.
FT SUSHI A1080.
FT SUSHI A1081.
FT SUSHI A1082.
FT SUSHI A1083.
FT SUSHI A1084.
FT SUSHI A1085.
FT SUSHI A1086.
FT SUSHI A1087.
FT SUSHI A1088.
FT SUSHI A1089.
FT SUSHI A1090.
FT SUSHI A1091.
FT SUSHI A1092.
FT SUSHI A1093.
FT SUSHI A1094.
FT SUSHI A1095.
FT SUSHI
```

	FT	DISULFID	1321	1370	BY SIMILARITY.
	FT	DISULFID	1350	1387	BY SIMILARITY.
	FT	DISULFID	1396	1439	BY SIMILARITY.
	FT	DISULFID	1426	1452	BY SIMILARITY.
	FT	DISULFID	1457	1498	BY SIMILARITY.
	FT	DISULFID	1484	1514	BY SIMILARITY.
	FT	DISULFID	1519	1568	BY SIMILARITY.
	FT	DISULFID	1548	1585	BY SIMILARITY.
	FT	DISULFID	1591	1633	BY SIMILARITY.
	FT	DISULFID	1619	1646	BY SIMILARITY.
	FT	DISULFID	1650	1693	BY SIMILARITY.
	FT	DISULFID	1679	1706	BY SIMILARITY.
	FT	DISULFID	1711	1753	BY SIMILARITY.
	FT	DISULFID	1739	1769	BY SIMILARITY.
	FT	DISULFID	1774	1823	BY SIMILARITY.
	FT	DISULFID	1803	1840	BY SIMILARITY.
	FT	DISULFID	1848	1891	BY SIMILARITY.
	FT	DISULFID	1877	1904	BY SIMILARITY.
	FT	DISULFID	1909	1952	BY SIMILARITY.
	FT	DISULFID	1938	1965	BY SIMILARITY.
	FT	CARBOHYD	56	56	N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHYD	252	252	N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHYD	410	410	N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHYD	447	447	N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHYD	509	509	N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHYD	578	578	N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHYD	702	702	N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHYD	860	860	N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHYD	897	897	N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHYD	959	959	N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHYD	1028	1028	N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHYD	1152	1152	N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHYD	1310	1310	N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHYD	1481	1481	N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHYD	1504	1504	N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHYD	1534	1534	N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHYD	1540	1540	N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHYD	1605	1605	N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHYD	1763	1763	N-LINKED (GLCNAC . .) (POTENTIAL).
	FT	CARBOHYD	1908	1908	N-LINKED (GLCNAC . .) (POTENTIAL).
SQ	SEQUENCE	2039 AA; 223589 MW; B2FD29B6AD3C5EB7 CRC64;			
Query Match		Best Local Similarity	17.8%;	Score 437; DB 1; Length 2039;	
		Matches 132; Conservative	27.7%; Pred. No. 1.6e-25;		
			63; Mismatches 149; Indels 132; Gaps		
OY	27	GQAATVYKCRPGRTLTGTVKCKNGEWPNSNRISICRKRKCGHGGPTPRGSRLAVGSEF	86		
DB	66	GTYLNTCECPGTSGRPFSITCLKNSTWGAOKDR--CRKSCSNRPDPVAGMYHIKG--I	121		
OY	87	BFGAKVVYTCDEGYOLLGEIDYRECDADG---WTNDPIICEVVK--LPVTELENGRIYS	141		
DB	122	QGSGIKVKSCTGRGHLIGS--SSATCIISDDVIWMNEPICDIRIPGGLPPT--ITNDGFIS	179		
OY	142	GAAPEDQETPCQVVRFECSNG-----FKISQXEMHCSEN-----GLMSNEKPQ-VKI	190		
DB	180	TNRR--NFHYGSVYTYYRNCPGSGRKAFELIGEPSIYCTSDDDQYGMSGAPAOIILN	236		
OY	191	SQLPRVENGG---DGITLKPVYKKENERQYCKOGFYVKKEGDVACTG--SGMNPQ-PSC	244		
DB	237	KCTPRNVENGIIIVSDN---RSLSFLINEVEVFHKQPFYVKKGPRAKYCOALNMWEPLPBC	293		
OY	245	EEMTCLTPYIPINGIYTPPHRIKHRIID---DETARYECKNGFYPATRSPVSKCTITYG-WIP-	298		
DB	294	SNV-QQPP--PYVLAHERQRDKDNFSPSQEVIFYISEEP-YDLRGAAASMRCTIPQGWMSA	349		
DB	350	AAPTCEVKSCDDEMGOLINLRV-----LFEPVNIOLGAKVDFVCEDEGFOLGKSSASY--	399		
OY	353	YLRCYVNG-----WEPEVPCLROCFIHFIYVYEGSSYQWRAYIBGQSAKVOCHSYSLPN--	406		
DB	400	---CYLAGESLIAMNSSVPVCEQ-----IFCPSPVPIPNKR	431		

Oy	407	-----	GODTYCTE-----	NG-W-SPPKC	424
D6	432	HTKPLEVPEPKAVNYTCDDPHDRGTSFDLIGESTINCTSDDPGNGVWSSPAPKC	1:                         		487
RESULT	8				
ID	CABP_MOUSE	STANDARD:	PRT:	469 AA.	
AC	P08607;				
DT	01-AUG-1988 (Rel. 08, Created)				
DT	01-FEB-1994 (Rel. 28, Last sequence update)				
DT	01-NOV-1997 (Rel. 35, Last annotation update)				
DE	C4B-BINDING PROTEIN PRECURSOR (C4BP).				
GN	C4BPA OR C4BP.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE: 88024997.				
RA	Kristensen T., Ogata R.T., Chung L.P., Reid K.B.M., Tack B.F.;				
RT	"cDNA structure of murine C4b-binding protein, a regulatory component				
RT	of the serum complement system.";				
RL	Biochemistry 26:4668-4674(1987).				
CC	-1- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT				
CC	ACTIVATION. IT BINDS AS A COROPCTOR TO C3B/C4B INACTIVATOR				
CC	(C3INA), WHICH THEN HYDROLIZES THE COMPLEMENT FRAGMENT C4B. IT				
CC	ALSO ACCELERATES THE DEGRADATION OF THE C4A/C2A COMPLEX (C3				
CC	CONVERTAS) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA				
CC	CHAIN BINDS C4B. IT INTERACTS ALSO WITH SERUM AMYLOID P COMPONENT.				
CC	-1- SUBUNIT: HOMOTETRAMER; NOT COVALENTLY LINKED. MOUSE LACKS THE				
CC	BETA CHAIN OF C4BP				
CC	-1- SIMILARITY: TO C4BP BETA CHAIN AND TO FIG APOLIPOPROTEIN R.				
CC	-1- SIMILARITY: CONTAINS 6 SUSHI (SCR) REPEATS.				
CC	-1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-44 IS THE INITIATOR.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration -				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	-----				
CC	EMBL: M17122; AAA37312.1; ALT_INTF.				
DR	PIR: A27117; NBMSC4.				
DR	HSSP: P10998; IYVC.				
DR	MGD: MG1:88229; C4BP.				
DR	INTERPRO: IPR000436; -.				
DR	PFAM: PF000084; sushi: 6.				
KW	Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal.				
FT	SIGNAL	1	56		
FT	CHAIN	57	469	C4B-BINDING PROTEIN.	
FT	DOMAIN	57	414	6 X SUSHI (SCR) REPEATS.	
FT	REPEAT	57	116	SUSHI 1.	
FT	REPEAT	119	177	SUSHI 2.	
FT	REPEAT	180	241	SUSHI 3.	
FT	REPEAT	244	300	SUSHI 4.	
FT	REPEAT	302	356	SUSHI 5.	
FT	REPEAT	358	414	SUSHI 6.	
FT	DISULFID	58	103	BY SIMILARITY.	
FT	DISULFID	88	115	BY SIMILARITY.	
FT	DISULFID	120	160	BY SIMILARITY.	
FT	DISULFID	146	176	BY SIMILARITY.	
FT	DISULFID	181	223	BY SIMILARITY.	
FT	DISULFID	209	240	BY SIMILARITY.	
FT	DISULFID	245	287	BY SIMILARITY.	
FT	DISULFID	273	299	BY SIMILARITY.	
FT	DISULFID	303	343	BY SIMILARITY.	
FT	DISULFID	329	355	BY SIMILARITY.	
FT	DISULFID	359	400	BY SIMILARITY.	
FT	DISULFID	386	413	BY SIMILARITY.	

SEQ	SEQUENCE	469 AA;	5151 MM;	41E137CB8D86C321 CRC64;
FT	CABOHYD	74	74	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CABOHYD	227	227	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CABOHYD	275	275	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CABOHYD	292	292	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CABOHYD	366	366	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CABOHYD	381	381	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CABOHYD	428	428	N-LINKED (GLCNAC. . .) (POTENTIAL).
SO	SEQUENCE	469 AA;	5151 MM;	41E137CB8D86C321 CRC64;

Query Match	15.7%;	Score 384.5;	DB 1;	Length 469;
Best Local Similarity	29.0%;	Pred. No. 3.1e-22;		
Matches 112; Conservative	64;	Mismatches 161;	Indels 49;	Gaps 21

[illegible]

RESULT	9			
CR2_MOUSE				
ID	CR2_MOUSE	STANDARD;	PRT;	1025 AA.
AC	P19070;			
DT	01-NOV-1990 (Rel. 16, Created)			
DT	01-NOV-1990 (Rel. 16, last sequence update)			
DT	01-FEB-1996 (Rel. 33, last annotation update)			
DE	COMPLEMENT RECEPTOR TYPE 2 PRECURSOR (CR2) (COMPLEMENT C3D RECEPTOR).			
GN	CR2.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BALB/C;			
RX	MEDLINE: 90229735.			
RA	Fingeroth J.D.;			
RT	"Comparative structure and evolution of murine CR2. The homolog of			
RT	the human C3d/iBbV receptor (CD21).";			
RL	J. Immunol. 144:3458-3467(1990).			
RL	[2]			
RP	SEQUENCE OF 12-1025 FROM N.A.			
RX	MEDLINE: 91010789.			
RA	Molina H., Kinoshita T., Inoue K., Carel J.C., Holers V.M.;			
RT	"A molecular and immunochemical characterization of mouse CR2.			
RT	Evidence for a single gene model of mouse complement receptors 1 and			
RL	2.";			
RL	J. Immunol. 145:2974-2983(1990).			
RL	[3]			
RP	SEQUENCE OF 343-401 AND 991-1025 FROM N.A.			

[illegible]

```

FT DISULFID 431 458 BY SIMILARITY.
FT DISULFID 463 501 BY SIMILARITY.
FT DISULFID 467 514 BY SIMILARITY.
FT DISULFID 519 568 BY SIMILARITY.
FT DISULFID 548 585 BY SIMILARITY.
FT DISULFID 594 636 BY SIMILARITY.
FT DISULFID 622 649 BY SIMILARITY.
FT DISULFID 654 689 BY SIMILARITY.
FT DISULFID 675 704 BY SIMILARITY.
FT DISULFID 709 752 BY SIMILARITY.
FT DISULFID 738 769 BY SIMILARITY.
FT DISULFID 778 820 BY SIMILARITY.
FT DISULFID 806 833 BY SIMILARITY.
FT DISULFID 841 884 BY SIMILARITY.
FT DISULFID 870 897 BY SIMILARITY.
FT DISULFID 902 945 BY SIMILARITY.
FT DISULFID 931 958 BY SIMILARITY.
FT CARBOHYD 77 77 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 113 113 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 276 276 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 316 316 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 364 364 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 380 380 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 484 484 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 527 527 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 615 615 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 639 639 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 694 694 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 754 754 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 790 790 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 813 813 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 823 823 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 851 851 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 901 901 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 989 921 YGS -> EFR (IN REF. 4).
FT CONFLICT 306 306 S -> T (IN REF. 2).
FT CONFLICT 520 520 P -> A (IN REF. 2).
FT CONFLICT 962 963 MISSING (IN REF. 4).
SQ SEQUENCE 1025 AA; 112994 MW; 19E518B9A0273694 CRC64;

```

Query Match 15.2%; Score 373; DB 1; Length 1025;  
 Best Local Similarity 25.7%; Pred. No. 5.5e-21;  
 Matches 139; Conservative 61; Mismatches 184; Indels 156; Gaps 35;

```

OY 3 CKGPP-----PRENSELSGKSEQLYSEGTQATYKCRPGYTLGTIVKVC-KNGEWPS 56
DB 402 COAPKIIINGOKEDSYLLN-----FDPGTSIRYSCDPGYLLVGEDTLHCTPECKWPI 454
OY 57 NPS-RICKRRPQC-HPGDTPGSGF-RLAVGSEFEFGAKVYVYCDGSGYLLGIDVRECA 113
DB 455 TPQCTVAECKPVPGLHFRPQNOQTRITAVNS-----SCDEGFO-TLSEAVQLCGG 503
OY 114 D-GWTDNDIPICEVVKCLPVTLENGRIYSGAEPDOEYFGGVFECSNG-----FKI 166
DB 504 TIPWIEIRLCKEIRCPPIVHNTHWSSSE---DVPYGVVYVIMCIPGEGEGRFL 560
OY 167 BQKMHCHSEN---GLMSNEKPPC---VETISCLPVRVNGDGYL--KPYKENERE 215
DB 561 IGEQTHCTSDSRGRGSSWSPAPLCKLSLPAVQCTDVHEN--GVKLTDNKAIFYNDVSV 618
OY 216 OYKCKGQVYKRGDAVCTGSG-WMPQ-----PSCDEN-----TC--- 249
DB 619 MFKCDGILSSGSSQIRCKANNTWPEKPLCKKEGCEPERVHGLPDDSHIKLVKRCQNG 678
OY 250 --LT-----PYIPGICITTPHRIKIRI--DDEIRYEC 276
DB 679 YQLTGYVEKCKNAENGWTFMKIEVCTVILGPPKIANGGHTGMAKHFLGNREVSYC 738
OY 277 KNGFPAIRSPYSKCTIT-----GWI-PAPRC---SLKPCFPPOKRGRLYYEESRPY 326
DB 739 DGGFYLLAEKSL-QCVNDSKSGHGSWSPPOCLOSSPLHCPDPPEYKHG----- 786

```

```

OY 327 FVPVIGREYS-----YYCDNGFTTPSQSYMDYLRG-TYNGWPEPEY-CLRCQIF-- 373
DB 787 --YKLKTHASRSHNDIVFVOCNGFIANGS---HLIRCHTNVTLPGVPTCIRKASLGC 841
OY 374 ---HYVEGESSYWO-RRYIEGOSAKVOCGSGYSLPFGDYYCTENG--WS-PPPKCVRI 427
DB 842 QSPSTIPNHNHGGSLARPPGMSWVSCYQGF-LMAGEARLICTHEGIVSQPPPCREYV 900

RESULT 10
CR2_HUMAN
ID CR2_HUMAN STANDARD; PRT; 1033 AA.
AC P20023;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE COMPLEMENT RECEPTOR TYPE 2 PRECURSOR (CR2) (COMPLEMENT C3D RECEPTOR)
DE (EPSTEIN-BARR VIRUS RECEPTOR) (CD21 ANTIGEN).
GN CR2 OR C3DR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OC [1]
RX SEQUENCE FROM N.A.
RA Fujisaku A., Harley J.B., Frank M.B., Gruner B.A., Frazier B.,
RA Holers V.M.;
RA "Genomic organization and polymorphisms of the human C3d/Epstein-Barr
RT virus receptor."
RL J. Biol. Chem. 264:2118-2125(1989).
RN [2]
RP SEQUENCE OF 226-233; 256-267; 332-341; 667-677 AND 898-908.
RX MEDLINE: 86287311.
RA Weis J.J., Fearon D.T., Klickstein L.B., Wong W.W., Richards S.A.,
RA de Bruyjn Kops A., Smith J.A., Weis J.H.;
RT "Identification of a partial cDNA clone for the C3d/Epstein-Barr
RT virus receptor of human B lymphocytes: homology to the receptor for
RT fragments C3b and C4b of the third and fourth components of
RT complement."
RT Proc. Natl. Acad. Sci. U.S.A. 83:5639-5643(1986).
RL CC -1- FUNCTION: RECEPTOR FOR COMPLEMENT C3D AND FOR THE EPSTEIN-BARR
CC VIRUS ON HUMAN B-CELLS. PARTICIPATES IN B LYMPHOCYTES ACTIVATION.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: MATURE B LYMPHOCYTES AND FOLLICULAR
CC DENDRITIC CELLS OF THE SPLEEN.
CC -1- SIMILARITY: TO MOUSE CR2. CD21 IS A MEMBER OF THE REGULATORS OF
CC COMPLEMENT ACTIVATION (RCA) FAMILY.
CC -1- SIMILARITY: CONTAINS 15 SUSHI (SCR) REPEATS.
CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD21 entry;
CC WWW="HTTP://WWW.NCBI.NLM.NIH.GOV/PROW/CD/CD21.HTM".
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC
DB EMBL: M26004; AAA35786.1; -
DB EMBL: M26016; AAB04638.1; -
DB EMBL: M24007; AAB04638.1; JOINED.
DB EMBL: M24008; AAB04638.1; JOINED.
DB EMBL: M24009; AAB04638.1; JOINED.
DB EMBL: M24010; AAB04638.1; JOINED.
DB EMBL: M24011; AAB04638.1; JOINED.
DB EMBL: M26009; AAB04638.1; JOINED.
DB EMBL: M26010; AAB04638.1; JOINED.
DB EMBL: M26011; AAB04638.1; JOINED.
DB EMBL: M26012; AAB04638.1; JOINED.
DB EMBL: M26013; AAB04638.1; JOINED.
DB EMBL: M26014; AAB04638.1; JOINED.
DB EMBL: M26015; AAB04638.1; JOINED.

```



FT	DISULFID	336	378	BY SIMILARITY.
FT	DISULFID	364	389	BY SIMILARITY.
FT	DISULFID	396	439	BY SIMILARITY.
FT	DISULFID	425	450	BY SIMILARITY.
FT	DISULFID	454	505	BY SIMILARITY.
FT	DISULFID	486	515	BY SIMILARITY.
FT	DISULFID	524	567	BY SIMILARITY.
FT	DISULFID	553	578	BY SIMILARITY.
FT	DISULFID	582	636	BY SIMILARITY.
FT	DISULFID	616	646	BY SIMILARITY.
FT	CARBONHD	162	162	N-LINKED (GLNAC. . .) (POTENTIAL).
FT	CARBONHD	545	545	N-LINKED (GLNAC. . .) (POTENTIAL).
FT	SITE	617	619	CELL ATTACHMENT SITE.
FT	VARIANT	450	450	C -> F (IN FL3B DEFICIENCY).
SO	SEQUENCE	661 AA;	75491 MW;	57A2FB4656085/F2 CRC64; /Frtid=VAR.007475.
Query Match                    13.7%; Score 336; DB 1; Length 661;				
Best Local Similarity         25.6%; Pred. No. 2,2e-18;				
Matches    118; Conservative   58; Mismatches   217; Indels    68; Gaps				
OY	1	EDCKPPPRENSELLSGWS--EOLYSEGTQATYKCRPYRTLG--TIYVKCNKGEM--	53	
Db	151	ETCLAP-----ELYNGVSTTQTTFKVKKDYQEACATGYTTAGCKRTEVEECITLYGMSL	204	
OY	54	VPSNPSTRICRRKPGCHPDPTFFGSFRL-----AVGSEFEFGAKYYVTODEGQLLGEI	106	
Db	205	TPK-----CTLKRC-----SSURLIENCFHPHRYKQTYEEGVVOFPCHENTYLSGS-	250	
OY	107	DYRECADGWNTDIPICE--VVKCLPYTELNGRIIVSAAEPDOEXYFGQVVFECNSGF	164	
Db	251	DLIOCYNGWVPESPVCGRNRKRPPPLPIINSKIQHST---TYNHGEIVHIECELNF	306	
OY	165	KIEGOKEMHSCENGLMSEKPOCEE---ISC-LRPVENGDGIYLKPVYKENREFQYKC	219	
Db	307	EIHGSAETR-C-EDGKW-T-EPRKCLEGGOKBACEEPEPLEMGANLHISKITIYNDDKYIYAC	364	
OY	220	KQGFVERKERDAVCTGSGWNPDPSCER--MTCL-TPYTIPNGIYTPHRI-KHRIDDEIRYE	275	
Db	365	KSGYLLHGSNEITCNRGKWTLPPECIVENENCKRHPIPVMMGNAVAADGLASAYANGSSVEYR	424	
OY	276	CKNPFYATRSRPVSKCTITTGITRAPROSLKPCDFPQRKHGRILYEESSRRPYFPYIGKEY	335	
Db	425	C-NETYLLRGSKISRCEDGKSSPPVC-LDEPTYNVMDYNNRNNIEMKMKKEGKVLHGDLI	482	
OY	336	SYCDNGFTTPSQSYWDYLRCITVGWMEPEVP-CLRQ-----CIFHYVEGESYSW	384	
Db	483	DVFCKGQGDLSPLLRPLSELVSQCNRGEVXKPYLCIRKSCKMCMSPLIRKIGVLIIST---	539	
OY	385	QRRRIEGOSAKVVOCHSGSIPLNODTYTYCTENGMSPPPKCV	425	
Db	540	VDIYENGSSVEYRCFDHHFLFGSRFA-YCLDGDMWTTPPLCL	579	
RESULT 12				
FL3B_MOUSE                  STANDARD;                  PRT;                  668 AA.				
AC	Q07968;			
DT	01-FEB-1995 (Rel. 31, Created)			
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	COAGULATION FACTOR XIII B CHAIN PRECURSOR (EC 2.3.2.13) (PROTEIN-			
DE	GLUTAMINE GAMMA-GLUTAMYLTRANSFERASE B CHAIN) (TRANSGLUTAMINASE B			
GN	FL3B OR CF13B.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;			
RN	Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-B10.D2/OSN; TISSUE=LIVER;			
IX	MEDLINE: 93224141.			



DR EMBL; M65292; AAA35946.1; -  
 DR EMBL; M65293; AAA35947.1; -  
 DR EMBL; X56209; CAA39666.1; -  
 DR PIR; A40455; A40455.  
 DR PIR; S14604; S14604.  
 DR HSP; P08603; 1HF1.  
 DR MIM; 134371; -  
 DR INTERPRO; IPR000436; -  
 DR PFAM; PF00084; sush1; 5.  
 DR Repeat; Glycoprotein, Sush1; Signal; Polymorphism.  
 FT SIGNAL 1 18  
 FT CHAIN 19 330  
 FT DOMAIN 22 328  
 FT REPEAT 22 84  
 FT REPEAT 86 141  
 FT REPEAT 146 202  
 FT REPEAT 207 263  
 FT REPEAT 265 328  
 FT CARBOHYD 126 126  
 FT CARBOHYD 194 194  
 FT VARIANT 157 157  
 FT VARIANT 159 159  
 FT VARIANT 175 175  
 FT VARIANT 175 175  
 FT CONFLICT 71 71  
 FT SEQUENCE 330 AA; 37661 MW; 8DCOD3F92A85E035 CRC64;

Query Match 12.8%; Score 315; DB 1; Length 330;  
 Best Local Similarity 43.2%; Pred. No. 4e-17;  
 Matches 51; Conservative 21; Mismatches 46; Indels 0; Gaps 0;

QY 307 CDPPEKNGRLYYEESRPYPVPKIKSEYSCDNGFTFSQSYWDYLRCTVNGMEPEVP 366  
 DB 23 CDEPKIHGILTYDEKPKPFQSVPTGEVYSCENFVSPKSPWTRITCTEEGMSPTPK 82  
 QY 367 CLNCGCFHHYVEGESSWQRRYIEGOSAKVOCHSGYSLPNCODTYCTENGMSPPRC 424  
 DB 83 CLNLCFEPVENGHSESSGQTHLEGDVITICMTGYRLQNNENNISCVERGWSPTPKC 140

## RESULT 14

VCP\_VACCV STANDARD; PRT; 263 AA.  
 ID P10998;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE COMPLEMENT CONTROL PROTEIN PRECURSOR (VCP) (SECRETORY PROTEIN 35)  
 DE (PROTEIN C3) (28 KDA PROTEIN).  
 GN C3L.  
 OS Vaccinia virus (strain WR), and Vaccinia virus (strain Copenhagen).  
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
 OC Orthopoxvirus.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 20-37.  
 RC STRAIN-WR;  
 RA MEDLINE; 88318974.  
 RA Kotwal G.J., Moss B.;  
 RT "Vaccinia virus encodes a secretory polypeptide structurally related  
 to complement control proteins.";  
 RL Nature 335:176-178(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-WR;  
 RX MEDLINE; 89073756.  
 RX Kotwal G.J., Moss B.;  
 RT "Analysis of a large cluster of nonessential genes deleted from a  
 vaccinia virus terminal transposition mutant.";  
 RL Virology 167:524-537(1988).  
 RN [3]

RP SEQUENCE FROM N.A.  
 RC STRAIN-COPENHAGEN;  
 RX MEDLINE; 91021027.  
 RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,  
 RA Paolletti E.;  
 RT "The complete DNA sequence of vaccinia virus.";  
 RL Virology 179:247-266(1990).  
 RN [4]  
 RP COMPLETE GENOME.  
 RC STRAIN-COPENHAGEN;  
 RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,  
 RA Paolletti E.;  
 RT Virology 179:517-563(1990).  
 RN [5]  
 RP FUNCTION.  
 RX MEDLINE; 92115714.  
 RA Isaacs S.N., Kotwal G.J., Moss B.;  
 RT "Vaccinia virus complement-control protein prevents  
 antibody-dependent complement-enhanced neutralization of infectivity  
 and contributes to virulence.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:628-632(1992).  
 RN [6]  
 RP STRUCTURE BY NMR OF 146-263.  
 RX MEDLINE; 97446168.  
 RA Wales A.P., Shaw G., Bright J., Perczel A., Campbell I.D.,  
 RA Barlow P.N.;  
 RT "NMR studies of a viral protein that mimics the regulators of  
 complement activation.";  
 RL J. Mol. Biol. 272:253-265(1997).  
 CC -1- FUNCTION: SERVES TO PROTECT THE VIRUS AGAINST COMPLEMENT ATTACK BY  
 INHIBITING BOTH CLASSICAL AND ALTERNATIVE PATHWAYS OF COMPLEMENT  
 ACTIVATION. BINDS C3B AND C4B.  
 CC -1- SIMILARITY: BELONGS TO THE SUPERFAMILY OF THE REGULATORS OF  
 COMPLEMENT ACTIVATION (RCA).  
 CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X13166; CAA31564.1; -  
 DR EMBL; M22812; AAA69605.1; -  
 DR EMBL; M35027; AAA47997.1; -  
 DR PIR; A31005; WMVZSP.  
 DR PDB; 1VVC; 03-DEC-97.  
 DR PDB; 1VVD; 03-DEC-97.  
 DR PDB; 1VVE; 03-DEC-97.  
 DR INTERPRO; IPR000436; -  
 DR PFAM; PF00084; sush1; 4.  
 KW Signal; Repeat; Sush1; 3D-structure.  
 FT SIGNAL 1 19  
 FT CHAIN 20 263  
 FT DOMAIN 20 262  
 FT REPEAT 20 82  
 FT REPEAT 85 144  
 FT REPEAT 147 202  
 FT REPEAT 205 262  
 FT DISULFID 21 70  
 FT DISULFID 54 81  
 FT DISULFID 86 126  
 FT DISULFID 112 143  
 FT DISULFID 148 190  
 FT DISULFID 176 201  
 FT DISULFID 206 248  
 FT DISULFID 234 261  
 FT SEQUENCE 263 AA; 28629 MW; E4322CC9A6EF8997 CRC64;

Query Match 12.8%; Score 314.5; DB 1; Length 263;







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2000, 16:48:55 ; Search time 70.54 Seconds

(without alignments)  
566.546 Million cell updates/sec

Title: US-09-316-163-14

Perfect score: 2454

Sequence: 1 EDCKGPPRENSILSGSMS.....DTYYCTENGWSPPCVRIK 428

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 297973 seqs, 93374136 residues

297973

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP:REMBL\_14:\*  
2: SP:archaea:\*  
3: SP:bacteria:\*  
4: SP:fungi:\*  
5: SP:human:\*  
6: SP:mammal:\*  
7: SP:mhc:\*  
8: SP:organelle:\*  
9: SP:phage:\*  
10: SP:plant:\*  
11: SP:rodent:\*  
12: SP:virus:\*  
13: SP:vertebrate:\*  
14: SP:unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1713	69.8	449	4	Q14570
2	1242	50.6	669	6	Q28085
3	813	33.1	452	11	Q61407
4	811	33.0	303	11	Q61405
5	569.5	23.2	1053	13	Q91275
6	566	23.1	808	11	Q61408
7	478	19.5	330	4	Q9UJ16
8	466	19.0	539	11	Q63135
9	453	18.5	679	11	Q99254
10	447.5	18.2	522	6	Q28769
11	447	18.2	2014	6	Q29530
12	437	17.8	559	4	Q9UOV2
13	437	17.8	2039	4	Q16745
14	437	17.8	2489	4	Q16744
15	436	17.8	1911	6	Q29528
16	434	17.4	661	6	Q29531
17	428	17.3	533	11	Q08569
18	424	17.3	645	12	Q9WRU2
19	414.5	16.9	522	6	Q28797

20	409	16.7	497	11	Q63612
21	396	16.1	579	11	Q60736
22	385.5	15.7	1045	6	Q46545
23	375	15.3	560	5	Q22328
24	373.5	15.2	974	5	P91658
25	371.5	15.1	1124	5	Q9VVR4
26	371	15.1	1087	4	Q14212
27	369.5	15.1	315	6	Q28770
28	369.5	15.1	1032	4	Q13866
29	346	14.1	657	4	Q14006
30	341	13.9	417	11	Q35520
31	332.5	13.5	331	4	Q92496
32	329	13.4	1652	5	Q9VU09
33	319	13.0	343	11	Q61406
34	318	13.0	483	11	Q64735
35	317	12.9	550	12	P88903
36	317	12.9	550	12	Q40912
37	315	12.8	330	4	Q9UJ17
38	314.5	12.8	259	12	P87616
39	307	12.5	243	4	Q14310
40	302.5	12.3	740	4	Q95508
41	298	12.1	754	6	Q28290
42	297.5	12.1	263	12	Q89859
43	297.5	12.1	263	12	Q07033
44	296.5	12.1	263	12	Q89076
45	296.5	12.1	285	6	Q19121

## ALIGNMENTS

RESULT 1  
ID Q14570 PRELIMINARY; PRT; 449 AA.  
AC Q14570; P78435;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
DE COMPLEMENT FACTOR H PRECURSOR.  
GN HF OR CFH.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE: 88134059.  
RT Ripoché J., Day A.J., Harris T.J.R., Sim R.B.;  
RL "The complete amino acid sequence of human complement factor H.";  
RN Biochem. J. 249:593-602(1988).  
RN [2]  
RP SEQUENCE OF 226-449 FROM N.A.  
RA MEDLINE: 86169701.  
RT Kristensen T., Wetsel R.A., Jack B.F.;  
RL "Structural analysis of human complement protein H: homology with Cab binding protein, beta 2-glycoprotein I, and the Ba fragment of B2.";  
RN J. Immunol. 136:3407-3411(1986).  
RN [3]  
RP SEQUENCE OF 1-19 FROM N.A.  
RA Vik D.P., Williams S.A.;  
RL Submitted (Apr-1996) to the EMBL/Genbank/DBJ databases.  
RN [4]  
RP SEQUENCE OF 1-9 FROM N.A.  
RA Dominguez O.;  
RL Theists (1993). Immunologia, Hospital Trias I Pujol, Spain.  
DR EMBL: X07523; CAA50403.1; -  
DR EMBL: M12383; AAB52013.1; -  
DR EMBL: U56979; AAB01987.1; -  
DR EMBL: Z29665; CAA82763.1; -  
DR HSSP: P10998; IYVC.  
DR INTERPRO: IPR000436; -  
DR PFM: PF00084; sush1; 7.  
KW SIGNAL.  
FT SIGNAL. 1 18 POTENTIAL.

FT CHAIN 19 449 POTENTIAL.  
SQ SEQUENCE 449 AA; 51033 MW; C2AD47F155343E3 CRC64;

Query Match 69.8%; Score 1713; DB 4; Length 449;  
Best Local Similarity 67.4%; Pred. No. 3.3e-142;  
Matches 288; Conservative 48; Mismatches 91; Indels 0; Gaps 0;

QY 1 EDCKGPPRENSSEILSGSNEQLYSEGTQATYKCRPGYRLTLGYKCKNGEWPNSPSR 60  
DB EDCNLELPRLRNETLILGMSDQTYPEGTOAIYKCRPGYRSLGNVIMYCKRGKGEVVALNPLR 78  
QY 61 ICRKRCGHPGDTPGRSFLAAGSEPEFGAKVYVTCDEGYQLLGEIDYRECDADGWTNDI 120  
DB 79 KQKRCGHPGDTPEGFTLTLGAVFEYGVKAVYTCNEGYQLLGEIDYRECDADGWTNDI 138  
QY 121 PICEVYKCLPVELENGRIYSGAEPDQYEFQVYFECNSGFKIEGOKEMHCSENGIM 180  
DB 139 PICEVYKCLPVTAPENKGIYSSAMEPDREYHGOAVRFPCNSGTYKIEGDEMHCSDDGFW 198  
QY 181 SNEKPCVEISCLPPRENGDGIYKLVYKEMERFQYKCKGQFVYKERGDVAVCTGSGWNP 240  
DB 199 SNEKPCVEISCKSPDIVNGSPISQKLIYKEMERFQYKCKMGYEYSERGDVAVCTESGWRP 258  
QY 241 QPSCEMTLTPITPGITTPHRIKIRIDELRYECKNGFYPATRSPVSKCTTGWITPAP 300  
DB 259 LPSCERKSCDNPYIPNGDYSPLRIKHRTGDEITYQCRNGEYPATRGNTACTSTGMIPAP 318  
QY 301 KCSLKRCDFPQFQHRLYEEESRRPYFPVPIGKEYSYCDNGFTTQSQSYWDYLACTVNG 360  
DB 319 KCTLKRCDFPDIKHGSLYHENMRPYFPVAVGVYTSYCDNEHTETSGSYWDHILHCTQDG 378  
QY 361 WEPEVCLRQCIFFHYVEGESSYWMRRYIEGOSAKVQCHSGYSLPNGODTYCTENGWSP 420  
DB 379 WSAVAVCLRKCYFPYLENGYNQNYGRKFVQGSIDVACHFGYALPRAQTLVTTCMENGWSP 438  
QY 421 PPKCVRI 427  
DB 439 TPKCIRV 445

RESULT 2  
028085 PRELIMINARY; PRT; 669 AA.  
AC 028085;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, last sequence update)  
DT 01-MAY-2000 (TREMblrel. 13, last annotation update)  
DE CCP MODULES 3-12, WITH PARTS OF CCP 2 AND 13 (FRAGMENT).  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-LIVER:  
RX MEDLINE; 96202005.  
RA Soames C.J., Day A.J., Sim R.B.;  
RT "Prediction from sequence comparisons of residues of factor H involved  
in the interaction with complement component C3b.";  
RT Blochem. J. 315:523-531(1996).  
DR EMBL; X98697; CAA6257.1; -  
DR HSSP; P10998; 1YVC.  
DR INTERPRO; IPR000436; -  
DR PFAM; PF00084; sush1; 11.  
FT NON\_TER 1  
FT TER 669  
SQ SEQUENCE 669 AA; 75683 MW; D0D9DB30E747AC2 CRC64;

Query Match 50.6%; Score 1242; DB 6; Length 669;  
Best Local Similarity 58.4%; Pred. No. 9.7e-101;  
Matches 206; Conservative 57; Mismatches 88; Indels 2; Gaps 1;

QY 76 GSFRLAVSGSEFEGAKVYVTCDEGYQLLGEIDYRECDADGWTNDIPICEVYKCLPTELE 135  
DB 3 GSFRLAEGNQFEGAKVYVTCDEGYQLLGEIDYRECDADGWTNDIPICEVYKCLPTELE 62  
QY 136 NGRVSGAEPDQYEFQVYFECNSGFKIEGOKEMHCSENGIMSNERNKPCQVEISCLPP 195  
DB 63 NGRVSGAEPDQYEFQVYFECNSGFKIEGOKEMHCSENGIMSNERNKPCQVEISCLPP 122  
QY 196 RVENGDIYKLVYKEMERFQYKCKGQFVYKERGDVAVCTGSGWNPQSCDEMTCLPTIP 255  
DB 123 VILNGQAVLPKATYKQNEVQYKCAAGFEYGOGRDVTCTKSGWTPATCTLEITCDPPRI 182  
QY 256 NGITTPHRIKIRIDELRYECKNGFYPATRSPVSKCTTGWITPAPCSKCPDFQFKHG 315  
DB 183 NGVIRPELSKRYGQDKITTECKKGFPEIRIGTDATCTRDGMYVPPCAMKPCSPYIKHG 242  
QY 316 RLYEESRRPYFPVPIGKEYSYCDNGFTTQSQSYWDYLACTVNGMEPEVPCLRQCIFFH 375  
DB 243 RLYE--SYRGYFARVNOQFVYSCDHHFPPRSQSRMDHLACTAEGMSPEEPCLRQCIFFH 300  
QY 376 VEYESSYWMRRYIEGOSAKVQCHSGYSLPNGODTYCTENGWSPPKCVRIK 428  
DB 301 LENGHNQHREKRYLQGETVAVHCYEGSYLQNDQNTMTCTESGMSPPRCIRVK 353

RESULT 3  
061407 PRELIMINARY; PRT; 452 AA.  
AC 061407;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, last sequence update)  
DT 01-MAY-2000 (TREMblrel. 13, last annotation update)  
DE COMPLEMENT FACTOR H-RELATED PROTEIN.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Vik D.P., Munoz-Canoves P., Kozono H., Martin L.G., Tack B.F.,  
RA Chaplin D.D.;  
RL J. Biol. Chem. 0:0-0(0).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Vik D.P., Munoz-Canoves P., Kozono H., Martin L.G., Tack B.F.,  
RL Submitted (Apr-1990) to the EMBL/GenBank/DBJ databases.  
DR EMBL; M29010; AAA37415.1; -  
DR HSSP; P08603; 1HFI.  
DR INTERPRO; IPR000436; -  
DR PFAM; PF00084; sush1; 7.  
SQ SEQUENCE 452 AA; 51602 MW; 2B697A4FFCE13CA CRC64;

Query Match 33.1%; Score 813; DB 11; Length 452;  
Best Local Similarity 71.0%; Pred. No. 2.3e-63;  
Matches 137; Conservative 18; Mismatches 38; Indels 0; Gaps 0;

QY 236 SCNMPQSCCEMTCLTYITNGITTPHRIKIRIDELRYECKNGFYPATRSPVSKCTIP 295  
DB 15 TAWLSTAKGEKTCSPYILNGITTPHRIKIRIDELRYECKNGFYPATRSPVSKCTIP 74  
QY 296 WIPAPRCSLKPCDFPQFQHRLYEEESRRPYFPVPIGKEYSYCDNGFTTQSQSYWDYL 355  
DB 75 WIPAPRCSLKPCDFPQFQHRLYEEESRRPYFPVPIGKEYSYCDNGFTTQSQSYWDYL 134  
QY 356 CTYNGMEPEVPCLRQCIFFHYVEGESSYWMRRYIEGOSAKVQCHSGYSLPNGODTYCTE 415  
DB 135 CTAGMEPEVPCLRQCIFFHYVEGESSYWMRRYIEGOSAKVQCHSGYSLPNGODTYCTE 194  
QY 416 NGMSPPPKCVRIK 428  
DB 195 NGMSPPPKCVRIK 207





"The murine complement receptor gene family. IV. Alternative splicing of Cr2 gene transcripts predicts two distinct gene products that share homologous domains with both human CR2 and CRI.".

12

SEQUENCE OF 21-367 FROM N.A.

RA MEDLINE: 95105691.

RA Kim Y.U., Kinoshita T., Molina H., Hourcade D., Seya T., Wagner L.M., Holers V.M.;

RT "Mouse complement regulatory protein Cr2/p65 uses the specific mechanisms of both human decay-accelerating factor and membrane cofactor protein.".

RL J. Exp. Med. 181:151-159(1995).

DR EMBL: U17126; AAA78271.1; JOINED.

DR EMBL: U17123; AAA78271.1; JOINED.

DR EMBL: U17124; AAA78271.1; JOINED.

DR EMBL: U17125; AAA78271.1; JOINED.

DR EMBL: U17126; AAA78271.1; JOINED.

DR EMBL: U17127; AAA78271.1; JOINED.

DR EMBL: M36470; AAA37449.1; JOINED.

DR HSSP: P10998; IYVC.

DR MGD: MGI:88489; Cr2.

DR INTERPRO: IPR000436; .

DR PFAM: PF00084; sushi; 10.

FT NON\_TER

FT SEQUENCE 679 AA; 74916 MW; 52FC0FDCE20CDC CRC64;

Query Match 18.5%; Score 453; DB 11; Length 679;

Best Local Similarity 28.6%; Pred. No. 1,3e-31;

Matches 126; Conservative 73; Mismatches 178; Indels 64; Gaps 27;

QY 24 YSEGTQATYKCRPGYRTLTGTVKCKNGEWPSPNSRICRKRPGCHPGDTPPGSFRLAVG 83

DB 42 FAIGTTWEKCPGPGYFRKSFITLTCTSKW--SDAQCFCKRKPCKMNPDEPHGSVHINTG 99

QY 84 SFEEGAVVYTCDEGYOLLGEIDYRECDAG--WTNDIPICEVVKCLPTELENGRIY 140

DB 100 --TEGSGTYYSCNNGYRLIGDSATCIVSDTWMNDMPLECESTPESPPAISNGDFY 157

QY 141 SGAAPDEYFEGQVAFECNSG-----FKIEGQEMHCSEN-----GLMSNEKPOCV-E 189

DB 158 SSSRD---SFGYGVVYTYCHGKMKREKFLDVGESKSTYCSKQNVGIMWSPPOCI 214

QY 190 ISCLPRVENG---DGIYKPYKENERPYQYCKOGFYKKEGDAVC--TSGMNPQ--PSC 244

DB 215 VKCPMPVEIENGIVESG--FKHSFPLNDVYIFKCKSGFTMKSRIMACOPNSWSPPLPTC 272

QY 245 EEMTCLTPY-IPNGIYTPHRIKRIHIDEIRYCKNGFYPATRSPVSKCTTG-WI-PAPR 301

DB 273 -FMGCLPPONILHGDYNNKKDEFFSYGOKVSTYCNPG-YTLIGTNLVECTSLGTSNIVPT 330

QY 302 CSLKPCD-FP-OFKHGRLYEESRRPYFPVPIGKEYSYCCNGFT--TPSQSYWDLRC 356

DB 331 CEVKSODALPNHLHGRVFLPNN-----LQLGAEVSFVCDLGLKGPSS-----QC 378

QY 357 TVNG---WEPEVPCLRQCTF---VYEGESSYQWRKRIEQAQKVKOCHSISLPNQD 409

DB 379 IPEGETVIMNKPVECEQISCDPPEVKNARKPYSLPIVGTALRYTCSPSYRL-IGEK 437

QY 410 TVYC-TEN---GM-SPPK 424

DB 438 AIFCISENVHATWDKAPPIC 458

RESULT 10

028769 PRELIMINARY: PRT: 522 AA.

AC 028769;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)

DE COMPLEMENT RECEPTOR (FRAGMENT).

OS Papio hamadryas (Hamadryas baboon).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;

OC Cercopithecinae; Papio.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-BONE MARROW;

RA Birmingham D.J., Logar C.M., Shen X.P., Chen W.;

RL Submitted (May-1996) to the EMBL/Genbank/DBJ databases.

DR EMBL: L77977; AAA99004.1; .

DR HSSP: P08603; IHFI.

DR INTERPRO: IPR000436; .

DR PFAM: PF00084; sushi; 7.

FT NON\_TER

FT SEQUENCE 522 AA; 56626 MW; 312FCBE03ADF19DC CRC64;

Query Match 18.2%; Score 447.5; DB 6; Length 522;

Best Local Similarity 29.7%; Pred. No. 2,9e-31;

Matches 134; Conservative 66; Mismatches 168; Indels 83; Gaps 31;

QY 27 GTQATYKCRPGYRTLTGTVKCKNGEWPSPNSRICRKRPGCHPGDTPPGSFRLAVGSEF 86

DB 71 GTYLKTECLPGYHGKPFSSILCNWTSADKR--CTRKSCKRNPDKPVNG--MVHVIKDI 126

QY 87 EFGAKVYVTCDEGYOLLGEIDYRECDAG--WTNDIPICEVVKCLPTELENGRIYS 141

DB 127 QFQSQIYNVSCNKGIRLIGS--SSATCIISGNTVIMDETPICETIIPGCLPPT-IANDFTIS 184

QY 142 GAAPEDEYF-FQGVAFECNSG-----FKIEGQEMHCSEN-----GLMSNEKPOCV-VE 189

DB 185 ---TSREYFPYGSVYVIRCNLISGRRKFLFELVGEPSYCTSKDDQVIGMSPAPCIIP 240

QY 190 ISCLPRVENGDIYL-KPYKENERPYQYCKOGFYKKEGDAVCIG--SGMNPQ--PSC 246

DB 241 NKCPMPVEIENGIVYSVRSLSFSLNEVEYFCQPFVWKGPRHQAQALNKWPELPSCSR 300

QY 247 MTCLTPYIPNGIYTPHRIKRIHIDEIRYCKNGFYPATRSPVSKCTTG-WI-APRCSL 304

DB 301 VCPPEPILGHEHTPSHQDPSFGQEVYSCPEG-YDLRGASALHCTPOGDMNEAIPCTY 359

QY 305 KPCD--FPQFKHGRLYEESRRPYFP--VPYIGKEYSYCCNGFTTPSQSYWDLRCTVNG 360

DB 360 KSCDPIGLQLEPHGRV-----LFLNLQGAKVSVCDEGRILGRF--ASHCVLAG 408

QY 361 ---WEPEVPCLRQCTF-----HY-VYEGESSYQWRKRIEQAQKVKOCHSISLPNQD 402

DB 409 MKALMNSVPCVEQ-IFCPNPAILNGRHIGAPLGDIPY-----GKEYSYICDHPDR 460

QY 403 SLP---NGDITYCTE---NG-W-SPPK 424

DB 461 GMTVNLIGESTIRCTSDPQNGWSSPAPK 491

RESULT 11

029530 PRELIMINARY: PRT: 2014 AA.

AC 029530;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)

DE COMPLEMENT RECEPTOR 1 (FRAGMENT).

GN CR1.

OS Pan troglodytes (Chimpanzee).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.

RN [1]

RP SEQUENCE FROM N.A.

RC MEDLINE: 94292799.

RA Birmingham D.J., Shen X.P., Hourcade D., Nickells M.W., Atkinson J.P.;

RT "Primary sequence of an alternatively spliced form of CR1. Candidate for the 75,000 M(r) complement receptor expressed on chimpanzee erythrocytes.".



DR EMBL; L17394; AAB60694.1; JOINED.  
DR EMBL; L17395; AAB60694.1; JOINED.  
DR EMBL; L17396; AAB60694.1; JOINED.  
DR EMBL; L17397; AAB60694.1; JOINED.  
DR EMBL; L17398; AAB60694.1; JOINED.  
DR EMBL; L17399; AAB60694.1; JOINED.  
DR EMBL; L17400; AAB60694.1; JOINED.  
DR EMBL; L17401; AAB60694.1; JOINED.  
DR EMBL; L17402; AAB60694.1; JOINED.  
DR EMBL; L17403; AAB60694.1; JOINED.  
DR EMBL; L17404; AAB60694.1; JOINED.  
DR EMBL; L17405; AAB60694.1; JOINED.  
DR EMBL; L17406; AAB60694.1; JOINED.  
DR EMBL; L17407; AAB60694.1; JOINED.  
DR EMBL; L17408; AAB60694.1; JOINED.  
DR EMBL; L17409; AAB60694.1; JOINED.  
DR EMBL; L17410; AAB60694.1; JOINED.  
DR EMBL; L17411; AAB60694.1; JOINED.  
DR EMBL; L17412; AAB60694.1; JOINED.  
DR EMBL; L17413; AAB60694.1; JOINED.  
DR EMBL; L17414; AAB60694.1; JOINED.  
DR EMBL; L17415; AAB60694.1; JOINED.  
DR EMBL; L17416; AAB60694.1; JOINED.  
DR EMBL; L17417; AAB60694.1; JOINED.  
DR EMBL; L17419; AAB60694.1; JOINED.  
DR EMBL; L17420; AAB60694.1; JOINED.  
DR EMBL; L17421; AAB60694.1; JOINED.  
DR EMBL; L17422; AAB60694.1; JOINED.  
DR EMBL; L17423; AAB60694.1; JOINED.  
DR HSSP; P08603; 1HFI.  
DR INTERPRO; IPR000436; -.  
DR INTERPRO; IPR000834; -.  
DR INTERPRO; IPR001424; -.  
DR PRAM; PR00084; sushi; 30.  
DR PROSITE; PS00087; SOD\_CU\_ZN\_1; UNKNOWN\_1.  
DR PROSITE; PS00133; CARBOXYPEPT\_ZN\_2; UNKNOWN\_2.  
SQ SEQUENCE 2039 AA; 223603 MM; B82FCB11C6B16635 CRC64;

## Query Match

Best Local Similarity 17.8%; Score 437; DB 4; Length 2039;  
Matches 132; Conservative 63; Mismatches 149; Indels 132; Gaps 31;

QY 27 GTGATYKCRPGYRTLGTIVVCKNGEVPNPSRICKRCPGCDPFGSFLAVSEF 86  
DB 66 GTLANECRGYSGRPSITCLKNSTWTKAKDR--CRKSCRNPPDVGNGAHYTKG--I 121  
QY 87 EFGAKVYVTCDEGYQLGELIDYRECDAG---WTNDIPICEVVK-LPYTELENGRIVS 141  
DB 122 QFGSQIKYSGTCKGYRLIGS--SSATCLISGDTYIMDNETPICDRIPGCLPPT-ITNGDFIS 179  
QY 142 GAAPROEYVEFGOVAFECGSG-----FKIEQKEMHCSEN-----GLMSNEKQC-VEI 190  
DB 180 TNRE---NFHYGSAVYTRKCPGSGGRKVFELVGEPSIYCTSNDOYGIWGPPOCIIPN 236  
QY 191 SCLPPEVENG---DSIYLKPYKENERFQYKCKGKGVYKERGDAVCTG-SGNWPO-PSC 244  
DB 237 KCPPEVENGILVSDN---RSLEFSLNEVYFRCOPGFVYMKGPGRVCKQALNKKPELPSC 293  
QY 245 EENATCLTPYIPNGIYPRIRIKRID---DEIRYECKNGFYPATRSPVSKCTITG-WIP 298  
DB 294 SRV-CQPP--PDVLAHERQORDKNFSPGGEVYFSCBP-YDLRGAASMHCTPQGDWSPA 349  
QY 299 APFGSLKPCD--FPOFKHGLYVESRPRFPY--PIGKITYSYCDNGFTT--PSOSYWD 352  
DB 350 APFCVKSCDDEFGQLNGKV-----LEPVNLQGAIVDEVCDEFGOLKSSASY-- 399  
QY 353 YLNCYVNG---WEPEYCLRCQIFHYVEYEGESSYQORRIEGLSAKVOCHSGSLPN-- 406  
DB 400 ---CVLAGHESLNNSSVPCQO-----IFCPSPVIVPDKR 431  
QY 407 -----GQDTYYCTE---NG-W-SPPPK 424

DB 432 HTGKPLEVFPFGKAVNYTCDDPHDRGTGFDLIGESTICTSDPGNGVWSSPAPRC 487  
RESULT 14  
ID 016744 PRELIMINARY; PRT: 2489 AA.  
AC 016744;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DE 01-MAY-2000 (Tremblrel. 13, Last annotation update)  
DE COMPLEMENT RECEPTOR 1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 94065175.  
RA Vik D.P., Wong W.W.;  
RT "Structure of the gene for the F allele of complement receptor type 1  
RT and sequence of the coding region unique to the S allele.";  
RL J. Immunol. 151:6214-6224(1993).  
[2]  
RP SEQUENCE FROM N.A.  
RA Vik D.P., Wong W.W.;  
RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.  
DR EMBL; L17418; AAB60695.1; -.  
DR EMBL; L17390; AAB60695.1; JOINED.  
DR EMBL; L17391; AAB60695.1; JOINED.  
DR EMBL; L17392; AAB60695.1; JOINED.  
DR EMBL; L17393; AAB60695.1; JOINED.  
DR EMBL; L17394; AAB60695.1; JOINED.  
DR EMBL; L17395; AAB60695.1; JOINED.  
DR EMBL; L17396; AAB60695.1; JOINED.  
DR EMBL; L17397; AAB60695.1; JOINED.  
DR EMBL; L17398; AAB60695.1; JOINED.  
DR EMBL; L17399; AAB60695.1; JOINED.  
DR EMBL; L17400; AAB60695.1; JOINED.  
DR EMBL; L17401; AAB60695.1; JOINED.  
DR EMBL; L17402; AAB60695.1; JOINED.  
DR EMBL; L17403; AAB60695.1; JOINED.  
DR EMBL; L17405; AAB60695.1; JOINED.  
DR EMBL; L17406; AAB60695.1; JOINED.  
DR EMBL; L17407; AAB60695.1; JOINED.  
DR EMBL; L17408; AAB60695.1; JOINED.  
DR EMBL; L17409; AAB60695.1; JOINED.  
DR EMBL; L17410; AAB60695.1; JOINED.  
DR EMBL; L17411; AAB60695.1; JOINED.  
DR EMBL; L17412; AAB60695.1; JOINED.  
DR EMBL; L17413; AAB60695.1; JOINED.  
DR EMBL; L17414; AAB60695.1; JOINED.  
DR EMBL; L17415; AAB60695.1; JOINED.  
DR EMBL; L17416; AAB60695.1; JOINED.  
DR EMBL; L17417; AAB60695.1; JOINED.  
DR EMBL; L17419; AAB60695.1; JOINED.  
DR EMBL; L17420; AAB60695.1; JOINED.  
DR EMBL; L17421; AAB60695.1; JOINED.  
DR EMBL; L17422; AAB60695.1; JOINED.  
DR EMBL; L17423; AAB60695.1; JOINED.  
DR EMBL; L17424; AAB60695.1; JOINED.  
DR EMBL; L17425; AAB60695.1; JOINED.  
DR EMBL; L17426; AAB60695.1; JOINED.  
DR EMBL; L17427; AAB60695.1; JOINED.  
DR EMBL; L17428; AAB60695.1; JOINED.  
DR EMBL; L17429; AAB60695.1; JOINED.  
DR EMBL; L17430; AAB60695.1; JOINED.  
DR HSSP; P08603; 1HFI.  
DR INTERPRO; IPR000436; -.  
DR INTERPRO; IPR000834; -.  
DR INTERPRO; IPR001424; -.  
DR PRAM; PR00084; sushi; 37.  
DR PROSITE; PS00087; SOD\_CU\_ZN\_1; UNKNOWN\_1.  
DR PROSITE; PS00133; CARBOXYPEPT\_ZN\_2; UNKNOWN\_3.  
SQ SEQUENCE 2489 AA; 272846 MM; CE11B53F2BAFAF6 CRC64;







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2000, 16:47:37 ; Search time 47.7 Seconds

(Without alignments)  
437.703 Million cell updates/sec

Title: US-09-316-163-11

Perfect score: 1876

Sequence: 1 EDNCNLPKRRMTTEILTGSMS.....PDIKHGCLYHMKRRPYPPV 329

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters: 182106

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: PIR65:\*  
2: PIR1:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1876	100.0	449	1 NBHHS	complement factor
2	1876	100.0	1231	1 NBHHS	complement factor
3	1261	67.2	1234	1 NBHSH	complement factor
4	903	48.1	669	2 S65551	factor H - bovine
5	486	25.9	1053	2 S46199	probable complemen
6	356	19.0	597	1 S53711	C4BP alpha chain p
7	353	18.8	597	1 NBHSC4	C4b-binding protei
8	353	18.8	676	2 A45300	complement C3b rec
9	350.5	18.7	2489	2 I73012	complement C3b/C4b
10	348	18.6	2014	2 I36936	complement recepto
11	343.5	18.3	482	2 A34924	complement C3b/C4b
12	337.5	18.0	560	2 T16833	complement C3b/C4b
13	335	17.9	360	2 T42821	hypothetical prote
14	323	17.2	497	2 JC2054	complement control
15	320.5	17.1	610	1 I46001	complement regulat
16	317	16.9	452	2 A35068	C4b-binding protei
17	317	16.9	558	2 S57953	complement factor
18	311.5	16.6	469	1 NBHSC4	C4BP protein alpha
19	307	16.4	433	2 A30550	C4b-binding protei
20	306	16.3	303	2 H35068	complement C3b/C4b
21	304	16.2	440	2 A43519	apolipoprotein H-I
22	303.5	16.2	1025	2 A43526	complement recepto
23	303	16.2	661	1 KFHU33	complement C3d/Eps
24	301.5	16.1	263	1 MNVZSP	coagulation factor
25	294	15.7	668	2 A46013	apolipoprotein H h
26	288	15.4	1091	2 PL0003	coagulation factor
27	286	15.2	830	2 A30359	complement C3d/Eps
28	285	15.2	579	2 A56740	p-selectin precurs
29	284.5	15.2	263	1 C36838	sperm-egg recognit
					complement control

30	284.5	15.2	263	2 T28450	hypothetical prote
31	283.5	15.1	263	2 B72152	B18L protein - var
32	279.5	14.9	345	1 NBMS	apolipoprotein H p
33	279	14.9	610	2 A35046	E-selectin precurs
34	276.5	14.7	612	2 S23174	endothelial leukoc
35	276	14.7	618	2 B42755	E-selectin precurs
36	274.5	14.6	349	2 G02913	sperm CD46 - human
37	274.5	14.6	369	2 I57998	membrane cofactor
38	274.5	14.6	768	2 A42755	p-selectin precurs
39	274	14.6	377	2 I54479	membrane cofactor
40	274	14.6	384	2 S01896	membrane cofactor
41	274	14.6	768	2 I53821	p-selectin rat
42	270.5	14.4	362	2 JC5194	membrane cofactor
43	270.5	14.4	369	2 JC5194	membrane cofactor
44	269.5	14.4	360	1 WMBE2E	membrane-bound com
45	267.5	14.3	302	1 WMBE1E	secretory compleme

#### ALIGNMENTS

```
RESULT 1
NBHHS
complement factor H precursor, short splice form - human
N:Alternate names: complement factor H-related protein; complement protein H
C:Species: Homo sapiens (man)
C>Date: 31-Dec-1993 #sequence: revision 23-Feb-1996 #text: change 22-Jun-1999
C:Accession: S03013; B60238; A27877; A61103; A26505; S10479
R:Ripoche, J.; Day, A.J.; Harris, T.J.R.; Slim, R.B.
Biochem. J. 249, 593-602, 1988
A:Title: The complete amino acid sequence of human complement factor H.
A:Reference number: S00254; MUID:88134059
A:Accession: S03013
A:Molecule type: mRNA
A:Residues: 1-449 <RIP>
A:Cross-references: EMBL:X07523; EMBL:Y00716; NID:g32492; PIRN:CA030403.1; PID:g75807
A:Note: part of this sequence, including the amino end of the mature protein was conf
A:Note: 402-Tyr was also found
R:Estaller, C.; Schwaebler, W.; Dierich, M.; Weiss, E.H.
Eur. J. Immunol. 21, 799-802, 1991
A:Title: Human complement factor H: two factor H proteins are derived from alternativ
A:Reference number: A60238; MUID:91184292
A:Accession: B60238
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-33;434-449 <EST>
A:Note: only portions of this 1.8 kilobase mRNA were sequenced
R:Schulz, T.F.; Schwaebler, W.; Stanley, K.K.; Weiss, E.; Dierich, M.P.
Eur. J. Immunol. 16, 1351-1355, 1986
A:Title: Human complement factor H: isolation of cDNA clones and partial cDNA sequenc
A:Reference number: A27877; MUID:87054207
A:Accession: A27877
A:Molecule type: mRNA
A:Residues: '1L',55-401,'Y',403-449 <SCH>
A:Note: an additional nucleotide present within the codon for Glu-310 was thought to
R:Schwaebler, W.; Zwirner, J.; Schult, T.F.; Linke, R.P.; Dierich, M.P.; Weiss, E.H.
Eur. J. Immunol. 17, 1485-1489, 1987
A:Title: Human complement factor H: expression of an additional truncated gene produc
A:Reference number: A61103; MUID:88055295
A:Accession: A61103
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 27-76 <SC2>
A:Note: this is a partial sequence of an alternatively spliced 1.8 kilobase mRNA that
R:Sim, R.B.; Discipio, R.G.
Biochem. J. 205, 285-293, 1982
A:Title: Purification and structural studies on the complement-system control protein
A:Reference number: A26505; MUID:83048213
A:Accession: A26505
A:Molecule type: protein
A:Residues: 19-20,'Q',22-29,'V',31-33,'Q',35 <SIM>
R:Barlow, P.N.; Norman, D.G.; Steinkasserer, A.; Horne, T.J.; Pearce, J.; Driscoll, P
Biochemistry 31, 3626-3634, 1992
```

A:Title: Solution structure of the fifth repeat of factor H: A second example of the com  
 A:Reference number: A44551; MUID:92232649  
 A:Contents: annotation; NMR structure determination, residues 264-292  
 R:Kristensen, T.; Wetzel, R.A.; Tack, B.F.  
 J. Immunol. 136, 3407-3411, 1986  
 A:Title: Structural analysis of human complement protein H: homology with C4b binding pr  
 A:Reference number: S10479; MUID:86169701  
 A:Accession: S10479  
 A:Molecule type: mRNA  
 A:Residues: 226-401, 'Y', 403-449 <KR1>  
 A:Cross-references: GB:M12383; NID:q180472; PIDN:AAA52013.1; PID:q180473  
 C:Comment: Factor H has also been found bound to cell membranes in an unknown manner. Hc  
 C:Comment: Alternative transcripts of 4.3, 1.8, and 1.4 kilobases are expressed in liver.  
 C:Genetics: <HF1>  
 A:Gene: GDB:HF1; HF  
 A:Cross-references: GDB:120041; OMIM:134370  
 A:Map position: 1q32-1q32  
 C:Genetics: <HF2>  
 A:Gene: GDB:HF2; HF  
 A:Cross-references: GDB:129095  
 A:Map position: 1q32-1q32  
 A:Note: The correspondence between the two loci and the sequences indicated is unclear;  
 C:Function:  
 A:Description: a cofactor in the inactivation of C3b by serine proteinase I; also increa  
 be alternative complement pathway  
 A:Pathway: complement alternate pathway  
 C:Superfamily: complement factor H; complement factor H repeat homology  
 C:Keywords: alternative splicing; complement alternate pathway; glycoprotein; plasma  
 F:1-18/Domain: signal sequence #status predicted <Sig>  
 F:19-449/Product: complement factor H, short splice form #status experimental <MAT>  
 F:21-80/Domain: complement factor H repeat homology <FH01>  
 F:85-141/Domain: complement factor H repeat homology <FH02>  
 F:146-205/Domain: complement factor H repeat homology <FH03>  
 F:210-262/Domain: complement factor H repeat homology <FH04>  
 F:246-248/Region: cell attachment (R-G-D) motif  
 F:267-320/Domain: complement factor H repeat homology <FH05>  
 F:325-385/Domain: complement factor H repeat homology <FH06>  
 F:389-442/Domain: complement factor H repeat homology <FH07>  
 F:21-66,52-80,85-129,114-141,146-192,178-205,210-251,237-262,267-309,294-320,325-374,357  
 F:217/Binding site: carbohydrate (Asn) (covalent) #status absent

Query Match 100.0%; Score 1876; DB 1; Length 449;  
 Best Local Similarity 100.0%; Pred. No. 3, 1e-127;  
 Matches 329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDCNELPRRNTETLLGWSQDQYPEGTOAIYKCRPGYRSLGNVIMCRKGEWALNPLR 60  
 |||||  
 DB 19 EDCNELPRRNTETLLGWSQDQYPEGTOAIYKCRPGYRSLGNVIMCRKGEWALNPLR 78  
 |||||  
 QY 61 KCKRPGGHGDPPECFETLLGQVNFYGVKAVYTCNEGVLGELINRREDVDGWTNDI 120  
 |||||  
 DB 79 KCKRPGGHGDPPECFETLLGQVNFYGVKAVYTCNEGVLGELINRREDVDGWTNDI 138  
 |||||  
 QY 121 PICEVAVKCLPVTAPENKIIYSASMEPDREYHFGQAVFVCSNGYKIGDEMHCSDDGFW 180  
 |||||  
 DB 139 PICEVAVKCLPVTAPENKIIYSASMEPDREYHFGQAVFVCSNGYKIGDEMHCSDDGFW 198  
 |||||  
 QY 181 SKKPKCVETISCKSPDYINSPIISQKIIYKENERFOYKCMNGYTSERGAVCIESGWRP 240  
 |||||  
 DB 199 SKKPKCVETISCKSPDYINSPIISQKIIYKENERFOYKCMNGYTSERGAVCIESGWRP 258  
 |||||  
 QY 241 LPSCSEKSCNPYIPNDYSPRLRIKHTGDEITVQCRNGFVPATRGMTAKTSGWIPAP 300  
 |||||  
 DB 259 LPSCSEKSCNPYIPNDYSPRLRIKHTGDEITVQCRNGFVPATRGMTAKTSGWIPAP 318  
 |||||  
 QY 301 RCTLKPCDYPDIKRGHGLYHENMRPPVPV 329  
 |||||  
 DB 319 RCTLKPCDYPDIKRGHGLYHENMRPPVPV 347

RESULT 2  
 NBIH0H

complement factor H precursor, long splice form - human  
 C:Species: Homo sapiens (man)  
 C:Date: 31-Dec-1993 #sequence-revision 31-Dec-1993 #text-change 22-Jun-1999  
 C:Accession: S00254; A60238; A54726; A61565; A26505; I72654; S66298  
 R:Ripoche, J.; Day, A.U.; Harris, T.J.R.; Sim, R.B.  
 Biochem. J. 249, 593-602, 1988  
 A:Title: The complete amino acid sequence of human complement factor H.  
 A:Reference number: S00254; MUID:88134059  
 A:Accession: S00254  
 A:Molecule type: mRNA  
 A:Residues: 1-1231 <RIP>  
 A:Cross-references: EMBL:Y00716; NID:q31964; PIDN:CAA68704.1; PID:q31965  
 A:Note: 402-Tyr was also found  
 A:Note: parts of this sequence, including the amino and carboxyl ends of the mature p  
 R:Estaller, C.; Schwaebler, W.; Dierich, M.; Weiss, E.H.  
 Eur. J. Immunol. 21, 799-802, 1991  
 A:Title: Human complement factor H: two factor H proteins are derived from alternat  
 A:Reference number: A60238; MUID:91184292  
 A:Accession: A60238  
 A>Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-56;1177-1231 <EST>  
 A:Note: only portions of this 4.3 kilobase mRNA were sequenced  
 R:Day, A.U.; Ripoche, J.; Lyons, A.; McIntosh, B.; Harris, T.J.R.; Sim, R.B.  
 Biosci. Rep. 7, 201-207, 1987  
 A:Title: Sequence analysis of a cDNA clone encoding the C-terminal end of human compl  
 A:Reference number: A54726; MUID:88025472  
 A:Accession: A54726  
 A>Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 'DFRN', 579-1231 <NAV>  
 A:Cross-references: GB:M17517; NID:q180497; PIDN:AAA52016.1; PID:q180498  
 A:Note: parts of this sequence were determined by protein sequencing  
 R:Ripoche, J.; Day, A.U.; Willis, A.C.; Belt, K.T.; Campbell, R.D.; Sim, R.B.  
 Biosci. Rep. 6, 65-72, 1986  
 A:Title: Partial characterization of human complement factor H by protein and cDNA se  
 A:Reference number: A61565; MUID:8618123  
 A:Accession: A61565  
 A>Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 'METGRNHLNAKI', 1050-1057, 'T', 1059-1102 <RI2>  
 R:Sim, R.B.; Discipio, R.G.  
 Biochem. J. 205, 285-293, 1982  
 A:Title: Purification and structural studies on the complement-system control protein  
 A:Reference number: A26505; MUID:83048213  
 A:Accession: A26505  
 A:Molecule type: protein  
 A:Residues: 19-20, 'Q', 22-29, 'V', 31-33, 'Q', 35 <SIM>  
 R:Barlow, P.N.; Norman, D.G.; Steinkasserer, A.; Horne, T.J.; Pearce, J.; Driscoll, P  
 Biochemistry 31, 3626-3634, 1992  
 A:Title: Solution structure of the fifth repeat of factor H: A second example of the  
 A:Reference number: A44551; MUID:92232649  
 A:Contents: annotation; NMR structure determination, residues 264-292  
 R:Norman, D.G.; Barlow, P.N.; Baron, M.; Day, A.J.; Sim, R.B.; Campbell, I.D.  
 J. Mol. Biol. 219, 717-725, 1991  
 A:Title: Three-dimensional structure of a complement control protein module in soluti  
 A:Reference number: A49224; MUID:91278097  
 A:Contents: annotation; NMR structure determination, residues 927-985  
 R:Estaller, C.; Koistinen, V.; Schwaebler, W.; Dierich, M.P.; Weiss, E.H.  
 J. Immunol. 146, 3190-3196, 1991  
 A:Title: Cloning of the 1.4-kb mRNA species of human complement factor H reveals a no  
 A:Reference number: I56100; MUID:91201892  
 A:Accession: I72654  
 A>Status: translated from GB/EMBL/DBD  
 A:Molecule type: mRNA  
 A:Residues: 1047-1231 <RES>  
 A:Cross-references: GB:M65294; NID:q183766; PIDN:AAA35948.1; PID:q183767  
 R:Carton, J.A.; Bates, R.C.; Smith, A.I.; Tetozy, T.; Arellano, A.; Gordon, D.L.; Burn  
 Biochim. Biophys. Acta 1289, 305-311, 1996  
 A:Title: Factor H co-purifies with thrombospondin isolated from platelet secretate.  
 A:Reference number: S66298; MUID:96205365  
 A:Accession: S66298

A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 411-419:574-578,580-582 <CAR>  
C:Comment: Factor H has also been found bound to cell membranes in an unknown manner. Hc  
C:Comment: Alternative transcripts of 4.3, 1.8, and 1.4 kilobases are expressed in liver.  
C:Genetics: <HF1>  
A:Gene: GDB:HF1; HF  
A:Cross-references: GDB:120041; OMIM:134370  
A:Map position: 1q32-1q32  
C:Genetics: <HF2>  
A:Gene: GDB:HF2; HF  
A:Cross-references: GDB:129095  
A:Map position: 1q32-1q32  
A:Note: the correspondence between the two loci and the sequences indicated is unclear.  
C:Function:  
A:Description: a cofactor in the inactivation of C3b by serine proteinase I; also increa  
he alternative complement pathway  
A:Pathway: complement alternate pathway  
C:Superfamily: complement factor H; complement factor H repeat homology  
C:Keywords: alternative splicing; complement alternate pathway; glycoprotein; plasma  
F:1-18/Domain: signal sequence #status predicted <Stc>  
F:19-1229/Product: complement factor H #status experimental <MPT>  
F:19-449/Product: complement factor H, short splice form #status experimental <MAT>  
F:21-80/Domain: complement factor H repeat homology <FH01>  
F:85-141/Domain: complement factor H repeat homology <FH02>  
F:146-205/Domain: complement factor H repeat homology <FH03>  
F:210-262/Domain: complement factor H repeat homology <FH04>  
F:246-320/Region: cell attachment (R-G-D) motif  
F:325-385/Domain: complement factor H repeat homology <FH05>  
F:389-442/Domain: complement factor H repeat homology <FH06>  
F:448-505/Domain: complement factor H repeat homology <FH07>  
F:509-564/Domain: complement factor H repeat homology <FH08>  
F:569-623/Domain: complement factor H repeat homology <FH09>  
F:630-684/Domain: complement factor H repeat homology <FH10>  
F:691-744/Domain: complement factor H repeat homology <FH11>  
F:753-803/Domain: complement factor H repeat homology <FH12>  
F:811-864/Domain: complement factor H repeat homology <FH13>  
F:870-926/Domain: complement factor H repeat homology <FH14>  
F:931-984/Domain: complement factor H repeat homology <FH15>  
F:989-1043/Domain: complement factor H repeat homology <FH16>  
F:1048-1102/Domain: complement factor H repeat homology <FH17>  
F:1109-1163/Domain: complement factor H repeat homology <FH18>  
F:1167-1228/Domain: complement factor H repeat homology <FH19>  
F:21-66-52-80-85-129-114-141-146-192-178-205-210-251-237-262-267-309-294-320-325-374-357  
F:1-803-811-853-839-864-870-915-901-926-931-973-959-984-989-1032-1018-1043-1048-1091-1077-  
F:217/Binding site: carbohydrate (Asn) (covalent) #status absent  
F:529-802-822-882-911/Binding site: carbohydrate (Asn) (covalent) #status experimental  
F:718-1029-1095/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 1876; DB 1; Length 1231;  
Best Local Similarity 100.0%; Pred. No. 9.2e-127;  
Matches 329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 EDCNELPPRRNTEITLGSMSDQTYEGTQAIYKCRPGYSLGNVIMVCKKGVALNPIR 60  
Db EDCNELPPRRNTEITLGSMSDQTYEGTQAIYKCRPGYSLGNVIMVCKKGVALNPIR 78  
QY 61 KCQRKPCGHGCTPGCTTLTGANFEYGVKAVYTCNBSGYQLLGEINRECDTGTNDI 120  
Db 79 KCQRKPCGHGCTPGCTTLTGANFEYGVKAVYTCNBSGYQLLGEINRECDTGTNDI 138  
QY 121 PICEVVKCLPVAPANGKIVSAMPDREYHFGAVRVCNSGGYITEGSEHNCSDDGFW 180  
Db 139 PICEVVKCLPVAPANGKIVSAMPDREYHFGAVRVCNSGGYITEGSEHNCSDDGFW 198  
QY 181 SKERKVCVEISCKSPDVINGSPIQKIYKENERPOYKCNMGYSESDAVCTESGWRP 240  
Db 199 SKERKVCVEISCKSPDVINGSPIQKIYKENERPOYKCNMGYSESDAVCTESGWRP 258  
QY 241 LPSCGEKCDNPIYINGDYSPLRIRHRRGDELTTCRNGGFYPATGNRAKCTSTGWIPAP 300  
|||||

Db 259 LPSCGEKCDNPIYINGDYSPLRIRHRRGDELTTCRNGGFYPATGNRAKCTSTGWIPAP 318  
QY 301 RCTLPCPDYDPDKHGHGLYHENMRPRFPV 329  
Db 319 RCTLPCPDYDPDKHGHGLYHENMRPRFPV 347

RESULT 3  
NBMSH  
complement factor H precursor - mouse  
N:Alternate names: protein beta-1-H  
C:Species: Mus musculus (house mouse)  
C:Date: 30-Sep-1987 #sequence, revision 30-Sep-1987 #text, change 22-Jun-1999  
C:Accession: A26154; 149711; 149728  
R:Kristensen, T.; Tack, B.F.  
Proc. Natl. Acad. Sci. U.S.A. 83, 3963-3967, 1986  
A:Title: Murine protein H is comprised of 20 repeating units, 61 amino acids in lengt  
A:Reference number: A26154; MUID:8623353  
A:Accession: A26154  
A:Molecule type: mRNA  
A:Residues: 1-1234 <KRI>  
A:Cross-references: GB:M12660; NID:q193724; PIDN:AAA37759.1; PID:q387181  
R:Natsune-Sakai, S.; Nonaka, M.; Nonaka, M.; Harada, Y.  
J. Immunol. 144, 358-362, 1990  
A:Title: Demonstration of an unusual allelic variation of mouse factor H by the compl  
A:Reference number: 149711; MUID:90111033  
A:Accession: 149711  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-18 <RR2>  
A:Cross-references: GB:M31979; NID:q193726; PIDN:AAA37762.1; PID:q193729  
R:Kunoz-Canoves, P.; Tack, B.F.; ViK, D.P.  
Biochemistry 28, 9891-9897, 1989  
A:Title: Analysis of complement factor H mRNA expression: Dexamethasone and IFN-gamma  
A:Reference number: 149728; MUID:90148935  
A:Accession: 149728  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-19 <RR2>  
A:Cross-references: GB:J02891; NID:q193805; PIDN:AAA37795.1; PID:q553926  
C:Comment: Two codominant alleles of factor H are present in mice.  
C:Comment: Factor H functions as a cofactor in the inactivation of C3b by serine prot  
(C5 convertase) in the alternative complement pathway.  
C:Genetics:  
A:Map position: 1  
C:Superfamily: complement factor H; complement factor H repeat homology  
C:Keywords: complement alternate pathway; duplication; glycoprotein; plasma  
F:1-18/Domain: signal sequence #status predicted <Stc>  
F:19-1234/Product: complement factor H #status predicted <MPT>  
F:21-80/Domain: complement factor H repeat homology <FH01>  
F:85-141/Domain: complement factor H repeat homology <FH02>  
F:146-205/Domain: complement factor H repeat homology <FH03>  
F:210-262/Domain: complement factor H repeat homology <FH04>  
F:246-320/Region: cell attachment (R-G-D) motif  
F:325-385/Domain: complement factor H repeat homology <FH05>  
F:389-442/Domain: complement factor H repeat homology <FH06>  
F:448-505/Domain: complement factor H repeat homology <FH07>  
F:509-564/Domain: complement factor H repeat homology <FH08>  
F:569-623/Domain: complement factor H repeat homology <FH09>  
F:629-683/Domain: complement factor H repeat homology <FH10>  
F:690-743/Domain: complement factor H repeat homology <FH11>  
F:752-803/Domain: complement factor H repeat homology <FH12>  
F:808-861/Domain: complement factor H repeat homology <FH13>  
F:867-931/Domain: complement factor H repeat homology <FH14>  
F:936-989/Domain: complement factor H repeat homology <FH15>  
F:994-1048/Domain: complement factor H repeat homology <FH16>  
F:1053-1107/Domain: complement factor H repeat homology <FH17>  
F:1114-1168/Domain: complement factor H repeat homology <FH18>  
F:1172-1233/Domain: complement factor H repeat homology <FH19>  
F:21-66-52-80-85-129-114-141-146-192-178-205-210-251-237-262-267-309-294-320-325-374-  
0-802-808-850-836-861-867-920-906-931-936-978-964-989-994-1037-1023-1048-1053-1096-10  
F:676-721-773-801-1030-1061-1225/Binding site: carbohydrate (Asn) (covalent) #status



RESULT 6  
 S53711  
 C4BP alpha chain precursor - rabbit  
 C:Species: Oryctolagus cuniculus (domestic rabbit)  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: S53711  
 R:de Frutos, P.G.; Dahlbaeck, B.  
 Biochim. Biophys. Acta 1261, 285-289, 1995  
 A:Title: cDNA structure of rabbit C4b-binding protein alpha-chain. Preserved sequence m  
 A:Reference number: S53711; MUID:95226458  
 A:Accession: S53711  
 A:Status: preliminary; nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-597 <DEF>  
 A:Cross-references: EMBL:Z55490  
 C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology  
 F:50-107/Domain: complement factor H repeat homology <FH1>  
 F:112-169/Domain: complement factor H repeat homology <FH2>  
 F:174-234/Domain: complement factor H repeat homology <FH3>  
 F:239-294/Domain: complement factor H repeat homology <FH4>  
 F:299-360/Domain: complement factor H repeat homology <FH5>  
 F:364-422/Domain: complement factor H repeat homology <FH6>  
 F:426-480/Domain: complement factor H repeat homology <FH7>  
 F:484-538/Domain: complement factor H repeat homology <FH8>

Query Match 19.0% Score 356; DB 1; Length 597;  
 Best local similarity 30.5%; Pred. No. 3e-18; Mismatches 157; Indels 38; Gaps 20;

Matches 106; Conservative 47; Mismatches 157; Indels 38; Gaps 20;  
 QY 2 DCNELPRRTEILLTGSWSQDTPESTQAIYKPGYRSIG- NVIMWC- KGEVALNPL 59  
 Db 49 DCGP- PPHLLFASSISELSENEYGTITLKTCRPGYTRNGLPILCKRGIM---SYD 104  
 QY 60 RRCQKRCPCGHPGTPGTFTLTGCVFEYGVKAVYTCNBSYQLLGELINRECDPTD---G 115  
 Db 105 TFCVKKRCRPGDGLPNGOYEVR--TDFSCQSQTFFSCSEYIIGSTT-SHCDIQEKGYE 161  
 QY 116 WTNDICEVYKCLPYTAPENCKIVSSAMBPDEHY-FGQAVFVNCSSGKIGIDERMHC 174  
 Db 162 WSPPLKCELVKCEPPNINNGHNG---NDIMHTGSSVYSCNPRSLGEMASTISC 217  
 QY 175 S---DGFWSKRRKPCVETLCKSPDVINGSPISQ-KIYKENERPOYKCNMGYSESG 229  
 Db 218 TVNKNVTGVWSPBPVCKEILICSPVNPVHGKIIISGFPIYNTKSDIMTYCIDDFVL--RG 275  
 QY 230 DAV--C-TESGRPLPS-CEEKSC-DNPYIPNG----DYSPLRIKHTGDEITTCRNG 279  
 Db 276 SSLIHCELDKSNMSPVCESSNCTGLPNVPVHASSOQGYQWSTKEGYVSGTEILRYKCRPG 335  
 QY 280 FYP-ATRGNTAKC-TSTGMIPARCTLKCDDYDIDKHGGIYHENMRP 325  
 Db 336 YRPVADPITVTCQEDLRMSPPFAGCAICCPHOLDNGAL-TEHSRNP 382

RESULT 7  
 NBHUC4  
 C4b-binding protein alpha chain precursor - human  
 N:Alternate names: C4BP; proline-rich protein  
 C:Species: Homo sapiens (man)  
 C:Date: 13-Aug-1986 #sequence\_revision 30-Jun-1993 #text\_change 22-Jun-1999  
 C:Accession: A33568; S02372; A90326; A24183; A93134; S29492; A31785; I52244; A03210  
 R:Matsumuchi, T.; Okamura, S.; Asano, T.; Sata, T.; Niho, Y.  
 Biochem. Biophys. Res. Commun. 165, 138-144, 1989  
 A:Title: Molecular cloning of the cDNA coding for proline-rich protein (PRP): identity c  
 A:Reference number: A33568; MUID:90073699  
 A:Accession: A33568  
 A:Molecule type: mRNA  
 A:Residues: 1-597 <MAI>  
 A:Cross-references: GB:M31452; NID:9190501; PIDN:AAA36506.1; PID:9190502  
 A:Note: the authors translated the codon GGA for residue 492 as Glu  
 R:Linlin, S.J.; Lewin, A.R.; Reid, K.B.M.  
 FEBS Lett. 232, 328-332, 1988

A:Title: Derivation of the sequence of the signal peptide in human C4b-binding protei  
 A:Reference number: S02372; MUID:86242821  
 A:Accession: S02372  
 A:Molecule type: mRNA  
 A:Residues: 17-81 <LI2>  
 A:Cross-references: EMBL:X07853  
 A:Note: although the sequence determined extends to residue 9 above, these authors as  
 R:Chung, L.P.; Bentley, D.R.; Reid, K.B.M.  
 Biochem. J. 230, 133-141, 1985  
 A:Title: Molecular cloning and characterization of the cDNA coding for C4b-binding pr  
 A:Reference number: A90326; MUID:86025405  
 A:Accession: A90326  
 A:Molecule type: mRNA  
 A:Residues: 80-597 <CH2>  
 A:Cross-references: GB:X02865; NID:929564; PIDN:CAA26617.1; PID:929565  
 A:Note: 92-Thr and 357-His were also found  
 R:Linlin, S.J.; Reid, K.B.M.  
 FEBS Lett. 204, 77-81, 1986  
 A:Title: Studies on the structure of the human C4b-binding protein gene.  
 A:Reference number: A24182; MUID:86301119  
 A:Accession: A24182  
 A:Molecule type: DNA  
 A:Residues: 203-288 <LIN>  
 A:Cross-references: EMBL:X04284; EMBL:X04296  
 R:Rodriguez de Cordoba, S.; Sanchez-Corral, P.; Rey-Campos, J.  
 J. Exp. Med. 173, 1073-1082, 1991  
 A:Title: Structure of the gene coding for the alpha polypeptide chain of the human co  
 A:Reference number: A43023; MUID:91217619  
 A:Contents: annotation; exon-intron boundaries  
 R:Chung, L.P.; Gagnon, J.; Reid, K.B.M.  
 Mol. Immunol. 22, 427-435, 1985  
 A:Title: Amino acid sequence studies of human C4b-binding protein: N-terminal sequenc  
 ced by cyanogen bromide treatment.  
 A:Reference number: A93134; MUID:85296001  
 A:Accession: A93134  
 A:Molecule type: protein  
 A:Residues: 49-81 <CH1>  
 A:Note: this paper reports amino-terminal sequences of the intact protein and of a nu  
 R:Hessing, M.; Kanters, D.; Takeya, H.; van't Veer, C.; Hackeng, T.M.; Iwanaga, S.; B  
 FEBS Lett. 317, 228-232, 1993  
 A:Title: The region Ser(333)-Arg(356) of the alpha-chain of human C4b-binding protein  
 A:Reference number: S29492; MUID:93146164  
 A:Accession: S29492  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 381-404 <HES>  
 R:Suzuki, K.; Nishio, J.  
 J. Biol. Chem. 263, 17034-17039, 1988  
 A:Title: Binding site for vitamin K-dependent protein S on complement C4b-binding pro  
 A:Reference number: A31785; MUID:89034204  
 A:Accession: A31785  
 A:Molecule type: protein  
 A:Residues: 495-505, 'X', 507-510, 'X', 512-515 <SUZ>  
 A:Note: this peptide appears to bind protein S  
 R:Dahlback, B.; Smith, C.A.; Muller-Eberhard, H.J.  
 Proc. Natl. Acad. Sci. U.S.A. 80, 3461-3465, 1983  
 A:Title: Visualization of human C4b-binding protein and its complexes with vitamin K-  
 A:Reference number: A93950; MUID:83221615  
 A:Contents: annotation; electron microscopy; three-dimensional structure; ligand bind  
 R:Asano, T.; Okamura, S.; Matsumuchi, T.; Sakamoto, N.; Sata, T.; Niho, Y.  
 Biochem. Biophys. Res. Commun. 174, 222-227, 1991  
 A:Title: Genomic organization of the alpha chain of the human C4b-binding protein gen  
 A:Reference number: I52244; MUID:91113159  
 A:Accession: I52244  
 A:Status: translation not shown; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-597 <ASO>  
 A:Cross-references: GB:M62486; NID:9190498; PIDN:AAA36506.1; PID:9190500  
 C:Comment: C4BP controls the classical pathway of complement activation. It binds as  
 the degradation of the C4bC2a complex (C3 convertase) by dissociating the complement  
 isofide bonds. The beta chain binds the vitamin K-dependent plasma protein S. Amino  
 actors V and VIII.

C:Comment: The molecule has a central body supporting seven tentacles (alpha chains), c

C:Genetics:

A:Gene: GDB:C4BPA

A:Cross-References: GDB:120568; OMIM:120830

A:Map position: 1q32-1q32

A:Introns: 48/1; 110/1; 143/2; 172/1; 236/1; 297/1; 362/1; 425/1; 482/1; 540/3

C:Complex: octamer of seven alpha chains and one beta chain

C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology

C:Keywords: acute phase; chylomicron; complement pathway; duplication; glycoprotein; pla

F:1-48/Domain: signal sequence #status predicted <SIS>

F:49-597/Product: C4b-binding protein alpha chain #status predicted <MAT>

F:50-108/Domain: complement factor H repeat homology <FH1>

F:113-170/Domain: complement factor H repeat homology <FH2>

F:175-234/Domain: complement factor H repeat homology <FH3>

F:239-294/Domain: complement factor H repeat homology <FH4>

F:299-360/Domain: complement factor H repeat homology <FH5>

F:364-422/Domain: complement factor H repeat homology <FH6>

F:381-404/Region: complement C4b binding #status predicted

F:426-480/Domain: complement factor H repeat homology <FH7>

F:484-538/Domain: complement factor H repeat homology <FH8>

F:521-506/Binding site: carbohydrate (asn) (covalent) #status experimental

Query Match 18.8%; Score 353; DB 1; Length 597;

Best Local Similarity 29.3%; Pred. No. 4.8e-18;

Matches 100; Conservative 50; Mismatches 141; Indels 50; Gaps 21;

20 SDQTYEGTQAIYKCRPGY-RSLGNVIMVCRK-GEWVALNPLRCKQRRPGHPDPEGT 77

119 --LRNQVEIKITDLSFGSQIEFSGSEGFLLIGSTSR-CEVQDRGVGWSHPLPQCEIVKC 175

129 LPTAENCGKIVSSAMPRDEYHFGQAVRVCNSGYKIEDEDEHHCSD-GEFWSKEK 184

176 KPRPDRNRNGHSGE---EMFYAVGFSVTVSCDPRFSLGHASISCEVENETIGVWRRSP 231

185 PKCEVSKSPDVINGSPISQ-KIYKENERFOYKCMGYEYSEKDAV--C-TESGWRP 240

232 PTECEIKTRKPRDYGSHGEMVSGFPIYNYKDTIVKCKQKQFVL--RGSSVICHADSKWNP 289

241 L-PSCEKSCSDN-PYIPNGDYS--PLRIK---HRTGDEITVOCNRGYPAT-RGNTAKC- 291

290 SPACPEPNSCINLPDIPHASWETYPPTKEDYVGVTLNRCHPGYKPTTDEPTVICO 349

292 TSTGWIAPRCTLKPCDYPDIKGGLYHENMRRP---YF 327

350 KNLRWTPYGGCEALCCEPEPLNNGEITQHRKSRANHCYF 390

RESULT 8

AA5900

complement C3b receptor type 2 long form precursor - mouse (fragment)

C:Species: Mus musculus (house mouse)

C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 08-Oct-1993

C:Accession: A45900; 148306

R:Kurtz, C.B.; O'Toole, E.; Christensen, S.M.; Weis, J.H.

J. Immunol. 144, 3581-3591, 1990

A:Title: The murine complement receptor gene family. IV. Alternative splicing of Cr2 ge

A:Reference number: A45900; M0ID:90229754

A:Accession: A45900

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-676 <KUR>

A:Cross-References: GB:M36470

A:Experimental sources: clone 31-1

R:Kim, Y.U.; Kinoshita, T.; Molina, H.; Hourcade, D.; Seya, T.; Wagner, L.M.; Holers, V.

J. Exp. Med. 181, 151-159, 1995

A:Title: Mouse complement regulatory protein Cr1/p65 uses the specific mechanisms of bc

A:Reference number: 148306; M0ID:95105691

A:Accession: 148306

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 21-367 <RES>

A:Cross-References: EMBL:U17128; NID:g9595980; PIDN:AAA74271.1; PID:g9595982

C:Genetics:

A:Gene: Cr2

A:Introns: 80/1; 113/2; 142/1; 275/1; 333/1

C:Superfamily: complement factor H repeat homology

F:22-78/Domain: complement factor H repeat homology <FH01>

F:83-140/Domain: complement factor H repeat homology <FH02>

F:145-211/Domain: complement factor H repeat homology <FH03>

F:217-272/Domain: complement factor H repeat homology <FH04>

F:276-331/Domain: complement factor H repeat homology <FH05>

F:336-394/Domain: complement factor H repeat homology <FH06>

F:399-458/Domain: complement factor H repeat homology <FH07>

F:467-523/Domain: complement factor H repeat homology <FH08>

F:531-587/Domain: complement factor H repeat homology <FH09>

F:592-648/Domain: complement factor H repeat homology <FH10>

Query Match 18.8%; Score 353; DB 2; Length 676;

Best Local Similarity 28.9%; Pred. No. 5.5e-18;

Matches 93; Conservative 61; Mismatches 136; Indels 32; Gaps 17;

20 SDQTYEGTQAIYKCRPGY-RSLGNVIMVCRK-GEWVALNPLRCKQRRPGHPDPEGT 78

37 SDKSEFAIGTTWEYKCRPGYFRKSFITTCLETSKW--SDAQOQCKRCPKMNPOEPLHGSV 94

79 TLTGAVFEXYKAVAVYCNQGYQLLEINREC---DTDGMTNDIPICEVVKCLPYAP 134

95 HINTG--IEFGSTITYSCNGYRLIGD-SSATGIVSDNTVMMDMDPLCSIPCESPAI 151

135 ENKIVSSAMPRDEYHFGQAVRVCNSG-----YIEDEDEHHCSD-GEFWSKEK 164

152 SDDFVSSSRD---SFEYGVAVVYTCHTGNREKLFQDVEKSTYCKSNQYGINMSP 208

185 PKCY-EISCKSPDVINGSPISQ-KIYKENERFOYKCMGYEYSEKDAV-C-TESGWRP 240

209 POCIPRYKCPMEIENLVSGFKHSFLLDYVIFKKSQFTMKGSRIAMCOPRKSMSPP 268

241 LPSCEKSCSDNYPINGDYSPLRIKHTGDEITVOCNRGYPATRGNTAKCTSTG-WI-P 298

269 LPFCFMCLPPONLHGDYNNKDEFFSVGQKVSYTCNDP-YTLIGTLVLCSTSLGTSWNT 327

299 APRCTLKPCD-YPD-IKHGGLY 318

328 VPTCEVKSQDAIPNHLHGRVF 349

RESULT 9

I73012

complement C3b/C4b receptor, membrane-bound form precursor - human

N:Alternate names: complement C3b/C4b receptor; complement receptor type 1 (CR1); sur

N:Contains: complement C3b/C4b receptor, secreted form

C:Species: Homo sapiens (man)

C:Date: 02-Jul-1996 #sequence\_revision 24-Nov-1999 #text\_change 31-Jan-2000

C:Accession: I73012; 156203; A47602; S03291; S03843; A28507; A24748; C24748

R:Vik, D.P.; Wong, W.W.

J. Immunol. 151, 6214-6224, 1993

A:Title: Structure of the gene for the F allele of complement receptor type 1 and seq

A:Reference number: I56203; M0ID:94065175

A:Accession: I73012

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-683, 'X', 685-1133, 'X', 1135-1471, 'X', 1473-2489 <VIK1>

A:Cross-References: GB:U17418; NID:g306678; PID:g451303

A:Accession: I56203

A:Status: translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-683, 'X', 685-994, 'A', 896-1000, 1451-1471, 'X', 1473-2489 <VIK2>

A:Cross-References: GB:U17418; NID:g306678; PID:g306660

R:Wong, W.W.; Cahill, J.M.; Rosen, M.D.; Kennedy, C.A.; Bonaccio, E.T.; Morris, M.J.;

J. Exp. Med. 169, 847-863, 1989

A:Title: Structure of the human CRI gene. Molecular basis of the structural and quantitative  
 A:Reference number: A47602; MUID:89176869  
 A:Accession: A47602  
 A:Molecule type: DNA  
 A:Residues: 1-41 <MON>  
 A:Cross-references: GB:X14893  
 R:Hourcade, D.; Miesner, D.R.; Atkinson, J.P.; Holers, V.M.  
 J. Exp. Med. 168, 1255-1270, 1988  
 A:Title: Identification of an alternative polyadenylation site in the human C3b/C4b repeat 1.  
 A:Reference number: S03291; MUID:89010527  
 A:Accession: S03291  
 A:Molecule type: mRNA  
 A:Residues: 26-584 <HOU>  
 A:Cross-references: EMBL:X14362; NID:g30197; PID:g736240  
 R:Klickstein, L.B.; Bartow, T.J.; Miletic, V.; Rabson, L.D.; Smith, J.A.; Fearon, D.T.  
 J. Exp. Med. 168, 1699-1717, 1988  
 A:Title: Identification of distinct C3b and C4b recognition sites in the human C3b/C4b repeat 1.  
 A:Reference number: S03843; MUID:89035992  
 A:Accession: S03843  
 A:Status: translation not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-894, 'A', 896-1000, 1451-2064, 'T', 2066-2276, 'P', 2278-2299, 'H', 2301-2325, 'T', 2327-2350, 'A', 2352-2376, 'G', 2378-2399, 'H', 2401-2425, 'T', 2427-2450, 'A', 2452-2476, 'G', 2478-2500, 'A', 2502-2526, 'T', 2528-2552, 'A', 2554-2578, 'G', 2580-2604, 'T', 2606-2630, 'A', 2632-2656, 'T', 2658-2682, 'A', 2684-2708, 'G', 2710-2734, 'T', 2736-2760, 'A', 2762-2786, 'G', 2788-2812, 'A', 2814-2838, 'T', 2840-2864, 'G', 2866-2890, 'A', 2892-2916, 'T', 2918-2942, 'A', 2944-2968, 'G', 2970-2994, 'T', 2996-3020, 'A', 3022-3046, 'T', 3048-3072, 'A', 3074-3098, 'T', 3100-3124, 'A', 3126-3150, 'T', 3152-3176, 'A', 3178-3202, 'T', 3204-3228, 'A', 3230-3254, 'T', 3256-3280, 'A', 3282-3306, 'T', 3308-3332, 'A', 3334-3358, 'T', 3360-3384, 'A', 3386-3410, 'T', 3412-3436, 'A', 3438-3462, 'T', 3464-3488, 'A', 3490-3514, 'T', 3516-3540, 'A', 3542-3566, 'T', 3568-3592, 'A', 3594-3618, 'T', 3620-3644, 'A', 3646-3670, 'T', 3672-3696, 'A', 3698-3722, 'T', 3724-3748, 'A', 3750-3774, 'T', 3776-3800, 'A', 3802-3826, 'T', 3828-3852, 'A', 3854-3878, 'T', 3880-3904, 'A', 3906-3930, 'T', 3932-3956, 'A', 3958-3982, 'T', 3984-4008, 'A', 4010-4034, 'T', 4036-4060, 'A', 4062-4086, 'T', 4088-4112, 'A', 4114-4138, 'T', 4140-4164, 'A', 4166-4190, 'T', 4192-4216, 'A', 4218-4242, 'T', 4244-4268, 'A', 4270-4294, 'T', 4296-4320, 'A', 4322-4346, 'T', 4348-4372, 'A', 4374-4398, 'T', 4400-4424, 'A', 4426-4450, 'T', 4452-4476, 'A', 4478-4502, 'T', 4504-4528, 'A', 4530-4554, 'T', 4556-4580, 'A', 4582-4606, 'T', 4608-4632, 'A', 4634-4658, 'T', 4660-4684, 'A', 4686-4710, 'T', 4712-4736, 'A', 4738-4762, 'T', 4764-4788, 'A', 4790-4814, 'T', 4816-4840, 'A', 4842-4866, 'T', 4868-4892, 'A', 4894-4918, 'T', 4920-4944, 'A', 4946-4970, 'T', 4972-4996, 'A', 4998-5022, 'T', 5024-5048, 'A', 5050-5074, 'T', 5076-5100, 'A', 5102-5126, 'T', 5128-5152, 'A', 5154-5178, 'T', 5180-5204, 'A', 5206-5230, 'T', 5232-5256, 'A', 5258-5282, 'T', 5284-5308, 'A', 5310-5334, 'T', 5336-5360, 'A', 5362-5386, 'T', 5388-5412, 'A', 5414-5438, 'T', 5440-5464, 'A', 5466-5490, 'T', 5492-5516, 'A', 5518-5542, 'T', 5544-5568, 'A', 5570-5594, 'T', 5596-5620, 'A', 5622-5646, 'T', 5648-5672, 'A', 5674-5698, 'T', 5700-5724, 'A', 5726-5750, 'T', 5752-5776, 'A', 5778-5802, 'T', 5804-5828, 'A', 5830-5854, 'T', 5856-5880, 'A', 5882-5906, 'T', 5908-5932, 'A', 5934-5958, 'T', 5960-5984, 'A', 5986-6010, 'T', 6012-6036, 'A', 6038-6062, 'T', 6064-6088, 'A', 6090-6114, 'T', 6116-6140, 'A', 6142-6166, 'T', 6168-6192, 'A', 6194-6218, 'T', 6220-6244, 'A', 6246-6270, 'T', 6272-6296, 'A', 6298-6322, 'T', 6324-6348, 'A', 6350-6374, 'T', 6376-6400, 'A', 6402-6426, 'T', 6428-6452, 'A', 6454-6478, 'T', 6480-6504, 'A', 6506-6530, 'T', 6532-6556, 'A', 6558-6582, 'T', 6584-6608, 'A', 6610-6634, 'T', 6636-6660, 'A', 6662-6686, 'T', 6688-6712, 'A', 6714-6738, 'T', 6740-6764, 'A', 6766-6790, 'T', 6792-6816, 'A', 6818-6842, 'T', 6844-6868, 'A', 6870-6894, 'T', 6896-6920, 'A', 6922-6946, 'T', 6948-6972, 'A', 6974-6998, 'T', 7000-7024, 'A', 7026-7050, 'T', 7052-7076, 'A', 7078-7102, 'T', 7104-7128, 'A', 7130-7154, 'T', 7156-7180, 'A', 7182-7206, 'T', 7208-7232, 'A', 7234-7258, 'T', 7260-7284, 'A', 7286-7310, 'T', 7312-7336, 'A', 7338-7362, 'T', 7364-7388, 'A', 7390-7414, 'T', 7416-7440, 'A', 7442-7466, 'T', 7468-7492, 'A', 7494-7518, 'T', 7520-7544, 'A', 7546-7570, 'T', 7572-7596, 'A', 7598-7622, 'T', 7624-7648, 'A', 7650-7674, 'T', 7676-7700, 'A', 7702-7726, 'T', 7728-7752, 'A', 7754-7778, 'T', 7780-7804, 'A', 7806-7830, 'T', 7832-7856, 'A', 7858-7882, 'T', 7884-7908, 'A', 7910-7934, 'T', 7936-7960, 'A', 7962-7986, 'T', 7988-8012, 'A', 8014-8038, 'T', 8040-8064, 'A', 8066-8090, 'T', 8092-8116, 'A', 8118-8142, 'T', 8144-8168, 'A', 8170-8194, 'T', 8196-8220, 'A', 8222-8246, 'T', 8248-8272, 'A', 8274-8298, 'T', 8300-8324, 'A', 8326-8350, 'T', 8352-8376, 'A', 8378-8402, 'T', 8404-8428, 'A', 8430-8454, 'T', 8456-8480, 'A', 8482-8506, 'T', 8508-8532, 'A', 8534-8558, 'T', 8560-8584, 'A', 8586-8610, 'T', 8612-8636, 'A', 8638-8662, 'T', 8664-8688, 'A', 8690-8714, 'T', 8716-8740, 'A', 8742-8766, 'T', 8768-8792, 'A', 8794-8818, 'T', 8820-8844, 'A', 8846-8870, 'T', 8872-8896, 'A', 8898-8922, 'T', 8924-8948, 'A', 8950-8974, 'T', 8976-9000, 'A', 9002-9026, 'T', 9028-9052, 'A', 9054-9078, 'T', 9080-9104, 'A', 9106-9130, 'T', 9132-9156, 'A', 9158-9182, 'T', 9184-9208, 'A', 9210-9234, 'T', 9236-9260, 'A', 9262-9286, 'T', 9288-9312, 'A', 9314-9338, 'T', 9340-9364, 'A', 9366-9390, 'T', 9392-9416, 'A', 9418-9442, 'T', 9444-9468, 'A', 9470-9494, 'T', 9496-9520, 'A', 9522-9546, 'T', 9548-9572, 'A', 9574-9598, 'T', 9600-9624, 'A', 9626-9650, 'T', 9652-9676, 'A', 9678-9702, 'T', 9704-9728, 'A', 9730-9754, 'T', 9756-9780, 'A', 9782-9806, 'T', 9808-9832, 'A', 9834-9858, 'T', 9860-9884, 'A', 9886-9910, 'T', 9912-9936, 'A', 9938-9962, 'T', 9964-9988, 'A', 9990-10014, 'T', 10016-10040, 'A', 10042-10066, 'T', 10068-10092, 'A', 10094-10118, 'T', 10120-10144, 'A', 10146-10170, 'T', 10172-10196, 'A', 10198-10222, 'T', 10224-10248, 'A', 10250-10274, 'T', 10276-10300, 'A', 10302-10326, 'T', 10328-10352, 'A', 10354-10378, 'T', 10380-10404, 'A', 10406-10430, 'T', 10432-10456, 'A', 10458-10482, 'T', 10484-10508, 'A', 10510-10534, 'T', 10536-10560, 'A', 10562-10586, 'T', 10588-10612, 'A', 10614-10638, 'T', 10640-10664, 'A', 10666-10690, 'T', 10692-10716, 'A', 10718-10742, 'T', 10744-10768, 'A', 10770-10794, 'T', 10796-10820, 'A', 10822-10846, 'T', 10848-10872, 'A', 10874-10898, 'T', 10900-10924, 'A', 10926-10950, 'T', 10952-10976, 'A', 10978-11002, 'T', 11004-11028, 'A', 11030-11054, 'T', 11056-11080, 'A', 11082-11106, 'T', 11108-11132, 'A', 11134-11158, 'T', 11160-11184, 'A', 11186-11210, 'T', 11212-11236, 'A', 11238-11262, 'T', 11264-11288, 'A', 11290-11314, 'T', 11316-11340, 'A', 11342-11366, 'T', 11368-11392, 'A', 11394-11418, 'T', 11420-11444, 'A', 11446-11470, 'T', 11472-11496, 'A', 11498-11522, 'T', 11524-11548, 'A', 11550-11574, 'T', 11576-11600, 'A', 11602-11626, 'T', 11628-11652, 'A', 11654-11678, 'T', 11680-11704, 'A', 11706-11730, 'T', 11732-11756, 'A', 11758-11782, 'T', 11784-11808, 'A', 11810-11834, 'T', 11836-11860, 'A', 11862-11886, 'T', 11888-11912, 'A', 11914-11938, 'T', 11940-11964, 'A', 11966-11990, 'T', 11992-12016, 'A', 12018-12042, 'T', 12044-12068, 'A', 12070-12094, 'T', 12096-12120, 'A', 12122-12146, 'T', 12148-12172, 'A', 12174-12198, 'T', 12200-12224, 'A', 12226-12250, 'T', 12252-12276, 'A', 12278-12302, 'T', 12304-12328, 'A', 12330-12354, 'T', 12356-12380, 'A', 12382-12406, 'T', 12408-12432, 'A', 12434-12458, 'T', 12460-12484, 'A', 12486-12510, 'T', 12512-12536, 'A', 12538-12562, 'T', 12564-12588, 'A', 12590-12614, 'T', 12616-12640, 'A', 12642-12666, 'T', 12668-12692, 'A', 12694-12718, 'T', 12720-12744, 'A', 12746-12770, 'T', 12772-12796, 'A', 12798-12822, 'T', 12824-12848, 'A', 12850-12874, 'T', 12876-12900, 'A', 12902-12926, 'T', 12928-12952, 'A', 12954-12978, 'T', 12980-13004, 'A', 13006-13030, 'T', 13032-13056, 'A', 13058-13082, 'T', 13084-13108, 'A', 13110-13134, 'T', 13136-13160, 'A', 13162-13186, 'T', 13188-13212, 'A', 13214-13238, 'T', 13240-13264, 'A', 13266-13290, 'T', 13292-13316, 'A', 13318-13342, 'T', 13344-13368, 'A', 13370-13394, 'T', 13396-13420, 'A', 13422-13446, 'T', 13448-13472, 'A', 13474-13498, 'T', 13500-13524, 'A', 13526-13550, 'T', 13552-13576, 'A', 13578-13602, 'T', 13604-13628, 'A', 13630-13654, 'T', 13656-13680, 'A', 13682-13706, 'T', 13708-13732, 'A', 13734-13758, 'T', 13760-13784, 'A', 13786-13810, 'T', 13812-13836, 'A', 13838-13862, 'T', 13864-13888, 'A', 13890-13914, 'T', 13916-13940, 'A', 13942-13966, 'T', 13968-13992, 'A', 13994-14018, 'T', 14020-14044, 'A', 14046-14070, 'T', 14072-14096, 'A', 14098-14122, 'T', 14124-14148, 'A', 14150-14174, 'T', 14176-14200, 'A', 14202-14226, 'T', 14228-14252, 'A', 14254-14278, 'T', 14280-14304, 'A', 14306-14330, 'T', 14332-14356, 'A', 14358-14382, 'T', 14384-14408, 'A', 14410-14434, 'T', 14436-14460, 'A', 14462-14486, 'T', 14488-14512, 'A', 14514-14538, 'T', 14540-14564, 'A', 14566-14590, 'T', 14592-14616, 'A', 14618-14642, 'T', 14644-14668, 'A', 14670-14694, 'T', 14696-14720, 'A', 14722-14746, 'T', 14748-14772, 'A', 14774-14798, 'T', 14800-14824, 'A', 14826-14850, 'T', 14852-14876, 'A', 14878-14902, 'T', 14904-14928, 'A', 14930-14954, 'T', 14956-14980, 'A', 14982-15006, 'T', 15008-15032, 'A', 15034-15058, 'T', 15060-15084, 'A', 15086-15110, 'T', 15112-15136, 'A', 15138-15162, 'T', 15164-15188, 'A', 15190-15214, 'T', 15216-15240, 'A', 15242-15266, 'T', 15268-15292, 'A', 15294-15318, 'T', 15320-15344, 'A', 15346-15370, 'T', 15372-15396, 'A', 15398-15422, 'T', 15424-15448, 'A', 15450-15474, 'T', 15476-15500, 'A', 15502-15526, 'T', 15528-15552, 'A', 15554-15578, 'T', 15580-15604, 'A', 15606-15630, 'T', 15632-15656, 'A', 15658-15682, 'T', 15684-15708, 'A', 15710-15734, 'T', 15736-15760, 'A', 15762-15786, 'T', 15788-15812, 'A', 15814-15838, 'T', 15840-15864, 'A', 15866-15890, 'T', 15892-15916, 'A', 15918-15942, 'T', 15944-15968, 'A', 15970-15994, 'T', 15996-16020, 'A', 16022-16046, 'T', 16048-16072, 'A', 16074-16098, 'T', 16100-16124, 'A', 16126-16150, 'T', 16152-16176, 'A', 16178-16202, 'T', 16204-16228, 'A', 16230-16254, 'T', 16256-16280, 'A', 16282-16306, 'T', 16308-16332, 'A', 16334-16358, 'T', 16360-16384, 'A', 16386-16410, 'T', 16412-16436, 'A', 16438-16462, 'T', 16464-16488, 'A', 16490-16514, 'T', 16516-16540, 'A', 16542-16566, 'T', 16568-16592, 'A', 16594-16618, 'T', 16620-16644, 'A', 16646-16670, 'T', 16672-16696, 'A', 16698-16722, 'T', 16724-16748, 'A', 16750-16774, 'T', 16776-16800, 'A', 16802-16826, 'T', 16828-16852, 'A', 16854-16878, 'T', 16880-16904, 'A', 16906-16930, 'T', 16932-16956, 'A', 16958-16982, 'T', 16984-17008, 'A', 17010-17034, 'T', 17036-17060, 'A', 17062-17086, 'T', 17088-17112, 'A', 17114-17138, 'T', 17140-17164, 'A', 17166-17190, 'T', 17192-17216, 'A', 17218-17242, 'T', 17244-17268, 'A', 17270-17294, 'T', 17296-17320, 'A', 17322-17346, 'T', 17348-17372, 'A', 17374-17398, 'T', 17400-17424, 'A', 17426-17450, 'T', 17452-17476, 'A', 17478-17502, 'T', 17504-17528, 'A', 17530-17554, 'T', 17556-17580, 'A', 17582-17606, 'T', 17608-17632, 'A', 17634-17658, 'T', 17660-17684, 'A', 17686-17710, 'T', 17712-17736, 'A', 17738-17762, 'T', 17764-17788, 'A', 17790-17814, 'T', 17816-17840, 'A', 17842-17866, 'T', 17868-17892, 'A', 17894-17918, 'T', 17920-17944, 'A', 17946-17970, 'T', 17972-17996, 'A', 18000-18024, 'T', 18026-18050, 'A', 18052-18076, 'T', 18078-18102, 'A', 18104-18128, 'T', 18130-18154, 'A', 18156-18180, 'T', 18182-18206, 'A', 18208-18232, 'T', 18234-18258, 'A', 18260-18284, 'T', 18286-18310, 'A', 18312-18336, 'T', 18338-18362, 'A', 18364-18388, 'T', 18390-18414, 'A', 18416-18440, 'T', 18442-18466, 'A', 18468-18492, 'T', 18494-18518, 'A', 18520-18544, 'T', 18546-18570, 'A', 18572-18596, 'T', 18598-18622, 'A', 18624-18648, 'T', 18650-18674, 'A', 18676-18700, 'T', 18702-18726, 'A', 18728-18752, 'T', 18754-18778, 'A', 18780-18804, 'T', 18806-18830, 'A', 18832-18856, 'T', 18858-18882, 'A', 18884-18908, 'T', 18910-18934, 'A', 18936-18960, 'T', 18962-18986, 'A', 18988-19012, 'T', 19014-19038, 'A', 19040-19064, 'T', 19066-19090, 'A', 19092-19116, 'T', 19118-19142, 'A', 19144-19168, 'T', 19170-19194, 'A', 19196-19220, 'T', 19222-19246, 'A', 19248-19272, 'T', 19274-19298, 'A', 19300-19324, 'T', 19326-19350, 'A', 19352-19376, 'T', 19378-19402, 'A', 19404-19428, 'T', 19430-19454, 'A', 19456-19480, 'T', 19482-19506, 'A', 19508-19532, 'T', 19534-19558, 'A', 19560-19584, 'T', 19586-19610, 'A', 19612-19636, 'T', 19638-19662, 'A', 19664-19688, 'T', 19690-19714, 'A', 19716-19740, 'T', 19742-19766, 'A', 19768-19792, 'T', 19794-19818, 'A', 19820-19844, 'T', 19846-19870, 'A', 19872-19896, 'T', 19898-19922, 'A', 19924-19948, 'T', 19950-19974, 'A', 19976-20000, 'T', 20002-20026, 'A', 20028-20052, 'T', 20054-20078, 'A', 20080-20104, 'T', 20106-20130, 'A', 20132-20156, 'T', 20158-20182, 'A', 20184-20208, 'T', 20210-20234, 'A', 20236-20260, 'T', 20262-20286, 'A', 20288-20312, 'T', 20314-20338, 'A', 20340-20364, 'T', 20366-20390, 'A', 20392-20416, 'T', 20418-20442, 'A', 20444-20468, 'T', 20470-20494, 'A', 20496-20520, 'T', 20522-20546, 'A', 20548-20572, 'T', 20574-20598, 'A', 20600-20624, 'T', 20626-20650, 'A', 20652-20676, 'T', 20678-20702, 'A', 20704-20728, 'T', 20730-20754, 'A', 20756-20780, 'T', 20782-20806, 'A', 20808-20832, 'T', 20834-20858, 'A', 20860-20884, 'T', 20886-20910, 'A', 20912-20936, 'T', 20938-20962, 'A', 20964-20988, 'T', 20990-21014, 'A', 21016-21040, 'T', 21042-21066, 'A', 21068-21092, 'T', 21094-21118, 'A', 21120-21144, 'T', 21146-21170, 'A', 21172-21196, 'T', 21198-21222, 'A', 21224-21248, 'T', 21250-21274, 'A', 21276-21300, 'T', 21302-21326, 'A', 21328-21352, 'T', 21354-21378, 'A', 21380-21404, 'T', 21406-21430, 'A', 21432-21456, 'T', 21458-21482, 'A', 21484-21508, 'T', 21510-21534, 'A', 21536-21560, 'T', 21562-21586, 'A', 21588-21612, 'T', 21614-21638, 'A', 21640-21664, 'T', 21666-21690, 'A', 21692-21716, 'T', 21718-21742, 'A', 21744-21768, 'T', 21770-21794, 'A', 21796-21820, 'T', 21822-21846, 'A', 21848-21872, 'T', 21874-21898, 'A', 21900-21924, 'T', 21926-21950, 'A', 21952-21976, 'T', 21978-22002, 'A', 22004-22028, 'T', 22030-22054, 'A', 22056-22080, 'T', 22082-22106, 'A', 22108-22132, 'T', 22134-22158, 'A', 22160-22184, 'T', 22186-22210, 'A', 22212-22

F:1684-1940/Domain:	complement factor H repeat	homology <FN19>
F:1873-1879/Domain:	complement factor H repeat	homology <FN18>
F:1749-1815/Domain:	complement factor H repeat	homology <FN17>
F:1655-1681/Domain:	complement factor H repeat	homology <FN16>
F:1432-1489/Domain:	complement factor H repeat	homology <FN15>
F:1236-1362/Domain:	complement factor H repeat	homology <FN14>
F:1172-1228/Domain:	complement factor H repeat	homology <FN13>
F:1041-1107/Domain:	complement factor H repeat	homology <FN11>
F:742-778/Domain:	complement factor H repeat	homology <FN10>
F:529-586/Domain:	complement factor H repeat	homology <FN9>
F:470-526/Domain:	complement factor H repeat	homology <FN8>
F:336-462/Domain:	complement factor H repeat	homology <FN7>
F:323-391/Domain:	complement factor H repeat	homology <FN6>
F:272-328/Domain:	complement factor H repeat	homology <FN5>
F:213-268/Domain:	complement factor H repeat	homology <FN4>
F:141-207/Domain:	complement factor H repeat	homology <FN3>
F:11-207/Domain:	complement factor H repeat	homology <FN0>

Query Match	18.6%;	Score 348;	DB 2;	Length 2014;
Best Local Similarity	28.4%;	Pred. No. 4.1e-17;		
Matches 96;	Conservative 57;	Mismatches 135;	Indels 50;	Gaps 23

Oy	20	SDFQYPRCTOAYIKRCRGYSLSLNTAVI-MV-CRKG-EWVAITPLRKOKPCKPHGTPGT	77
		: : :   :   :     :   :	
Dd	1387	NDFEFPYGTSLANTCGRGY--FGKMFSTOLENLMWSVED--NCRRSCGPPE-PENG	1441
Oy	78	FLLTGNVFGEYKAAYTTCMEGTYQLLGELINRECDTDG----WTNDIPCEYVKCLPYTA	133
		: : :   :   :     :   :	
Dd	1442	MVHINTDPT-QGSGTVANSCSEGFRLIGSPS-TTCLIVSGNNVTWDKKAPICELISEPPT	1499
Oy	134	PENCGVISSAEPREDREHFQOAVFYCNCS-----YKTEGDZEMHCs--DD--GEWSKE	183
		: :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Dd	1500	ISNGDEFSNNR---ASPHNGIVYTQCHTPGDEOLFETLYGSRSLYCTSKDDQGVWMSPP	1556
Oy	184	KPKCYEIS-CKSPVINGSPI-SQKITTKNEFQFKCMWGVEYSBRGDVATSESG-RRP	240
		: :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Dd	1557	PPRCISTNKCTAPAEVENAIRPNGRSFFSLTELVRRCOPGGVMWGSHTVOQTNGRMGP	1616
Oy	241	-LPSCEEKSCNPITYPMGDSDPSULIKHR-----TGDEITYOCRNGEYPATRIGNTAICTSG	295
		: :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Dd	1617	KLPHCSRVCQDPPELLGHEHP--SHQDNFSGGOEVFYSCERG-YDLRGASLHCTPOG	1672
Oy	236	-WIP-APRCTLKPD--YPRDKHGGLHEMMRRPYEPY 329	
		:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Dd	1673	DMSPEARCTIVASCDDDELGLQLPBG-----RVLEPFL 1702	

```

RESULT 11
A34924
complement C3b/C4b receptor-like protein precursor - human
C:Species: Homo sapiens (man)
C:Date: 27-Jul-1990 #sequence_revision 11-Apr-1997 #text_change 03-Dec-1999
C:Accession: A34924; S03292
R:Houcard, D.; Miesner, D.R.; Bee, C.; Zeldes, W.; Atkinson, J.P.
J. Biol. Chem. 265, 974-980, 1990
A:Title: Duplication and divergence of the amino-terminal coding region of the complement
A:Reference number: A34924; MUID:90110163
A:Accession: A34924
A:Molecule type: DNA
A:Residues: 1-479 <HOU>
A:Cross-references: GB:J05195
R:Houcard, D.; Miesner, D.R.; Atkinson, J.P.; Holers, V.M.
J. Exp. Med. 168, 1255-1270, 1988
A:Title: Identification of an alternative polyadenylation site in the human C3b/C4b receptor
type 1.
A:Reference number: S03291; MUID:89010527

```

C:Superfamily: Cab-binding protein alpha chain, complement factor H repeat homology  
E:43-99/Domain: complement factor H repeat homology <FH01>  
F:104-161/Domain: complement factor H repeat homology <FH02>  
F:166-232/Domain: complement factor H repeat homology <FH03>  
F:238-293/Domain: complement factor H repeat homology <FH04>  
F:297-353/Domain: complement factor H repeat homology <FH05>  
F:358-416/Domain: complement factor H repeat homology <FH06>  
F:421-478/Domain: complement factor H repeat homology <FH07>

Query Match	18.3%	Score 343.5;	DB 2;	Length 482;
Best Local Similarity	30.1%;	Pred. No. 1.8e-17;		
Matches 104;	Conservative 52;	Mismatches 126;	Indels 63;	Gaps 25;

[illegible]

RESULT 12  
T16833  
hypothetical protein T07H6.5 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 17-Mar-2000  
C:Accession: T16833  
R:Geisels, C.  
submitted to the EMBL Data Library, April 1996  
A:Description: The sequence of C. elegans cosmid T07H6.  
A:Reference number: Z18586  
A:Accession: T16833  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residue: 1-350 <GET>  
A:Cross-References: EMBL:U53344; NID:g1253886; PID:g1253889; PIDN:AAA96225.1; GSPDB:G  
A:Experimental source: strain Bristol NZ; clone T07H6  
C:Genetics:  
A:Gene: CESP:T07H6.5  
A:Map position: X  
A:Introns: 14/1: 75/1; 102/3: 128/1; 186/1: 272/2; 326/1: 361/1; 422/1: 475/1; 527/1  
C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology

```

Query Match 18.0%; Score 337.5; DB 2; Length 560;
Best Local Similarity 23.0%; Pred. No. 5.8e-17;
Matches 100; Conservative 57; Mismatches 142; Indels 135; Gaps

QY 3 CNELPRTRETELTGWSMDQYTPPESTQAIYKCRPQYSLGNVINCRRG--EWVALNDL-- 59
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 17 CHPLPIENGFIIRA--GDLPF--GSNAEYGCNKGYILVGAASQRCQANKNEMSSSQYVCR 72
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 60 -----RKQKRPFG 68
   : | : |

Db 73 LQIKCPPEPEIFPVAHVDGSSFGSEIDLNAEGISTSKCLLNKNVAQMPGDLRCKANACP 122

```

QY 69 HPGDPTGFTLTGGNVEYGVKAVYTCNBSGOLLGELINRECDTGD-WTNDIPICEVVK 127  
 Db 133 DRGIENG---LRGDDFFEPHVKYSCNPGFLVGST--RQSSNGEMTEPAPCKATE 188  
 QY 128 CLPYTABENGKIVSSAMEPDRREYHFGQAVRFVCSNGYKIEGDEEMHCDDGFWSEKREKC 187  
 Db 189 CSRPSSPLHGKVGSSSLT-----YQSVVYVSCDHGRVLGQVORICLAEIKNGENRRC 242  
 QY 188 VEISCK-SPDYING-----SPT----- 203  
 Db 243 EELRCVSPLPLPNCYIEGSETSFGAVAVFRCLTETMTEGASKAKMEDQWSPAPICLA 302  
 QY 204 SOKIYKENERFOYK-----CNMGXEYSEKADACTESGMRPLPSCERKS 248  
 Db 303 SCRPVHQNKGIRKSKSGQLASSKAYIECNKOHANIDERLICSNTSMHVPVCSPLS 362  
 QY 249 CDN--PYIPNGDYSPLRIKHTGDEITYQCRNGFPYATRGNTAKTSTGWT---PAPRCT 303  
 Db 363 CHMWPPEVPHARI--LFSKSHSGSLAKYECNNGYHPRNNQIICLXGEWTKDGPMPKCL 420  
 QY 304 LKPCDVPDIXHGGI 317  
 Db 421 PSCWCHPSKTYGTL 434

# RESULT 13 T42921

complement control protein homolog ccph - ateline herpesvirus 3 (strain 73)  
 C:Species: ateline herpesvirus 3  
 A:Variety: strain 73  
 C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 04-Mar-2000  
 A:Accession: T42921  
 R:Albrecht, J.C.; Fleckenstein, B.  
 Submitted to the EMBL Data Library, August 1998  
 A:Description: Primary structure of the herpesvirus ateles genome.  
 A:Reference number: Z22274  
 A:Accession: T42921  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-360 <ALB>  
 A:Cross-references: EMBL:AF083424; PIDN:MAC95530.1  
 A:Experimental source: strain 73  
 C:Superfamily: herpesvirus complement control protein; complement factor H repeat homolo

Query Match 17.9%; Score 335; DB 2; Length 360;  
 Best Local Similarity 33.4%; Pred. No. 5.5e-17;  
 Matches 103; Conservative 32; Mismatches 119; Indels 54; Gaps 20;

QY 8 PRNRTEI-----LNGSWSDQTYPEGTQAIYKCRPGYRSLGNVT--MVCRRKGFWALNPL 59  
 Db 24 PRNRVYSLRYVNIN--SSGSYPNCTITLOVYCRKCY--IGROIQVTVYVNGWVTPN-- 77  
 QY 60 RKCQKRPCGHPDTPPGTTLTGGNVEYGVKAVYTCNBSGOLLGELINREC--DTDG-- 115  
 Db 78 -RCQKRCRSTPADLNGWYTVT--GNLY-YGSVITYTCNFGYQLGSPF--SSCLLGGDGRV 133  
 QY 116 -WTNDIPICEVVKCLPYTABENGKIVSSAMEPDRREYH-FGQAVRFVCSNGYK--IGDDEE 171  
 Db 134 NMTTPRPICEITKCKRPPPIANGTHTNT-----KEYITYLDAVATYSCNDETKTLTLGPS 188  
 QY 172 MHCSDDGFW-SREKPRCWEISCKSPDYINGSPISOKIYKENERFOY--KCNMGXEYSE 227  
 Db 189 KOCSETEGRVAPDEETKCEKFKVCKIPQVANGHVEVKT--SNWQYQYVAMIKCDKGRLOG 246  
 QY 228 RCDVACTESGMRP-LPSCERKSCNDPYINGDYSPLRIKHTGDEITYQCRNGFPYATRG 286  
 Db 247 ETPNCKNGWMPALPTCK-----PAPRGD-----MPHIDSGEDTST-----PSGRN 290  
 QY 287 NTAKTST 294  
 Db 291 CNQNCST 298

## RESULT 14 JC2054

complement regulatory protein, 512 antigen precursor - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 27-Jun-1994 #sequence\_revision 27-Jun-1994 #text\_change 21-Jan-2000  
 A:Accession: JC2054; PC2027  
 R:Sakurada, C.; Seno, H.; Dohl, N.; Takizawa, H.; Nonaka, M.; Okada, N.; Okada, H.  
 Biochem. Biophys. Res. Commun. 198, 819-826, 1994  
 A:Title: Molecular cloning of the rat complement regulatory protein, 512 antigen.  
 A:Reference number: JC2054; MUID:94161746  
 A:Accession: JC2054  
 A:Molecule type: mRNA  
 A:Residues: 1-497 <SAK>  
 A:Accession: PC2027  
 A:Molecule type: protein  
 A:Residues: 39-51 <SN2>  
 C:Comment: This protein plays a critical role in protection against complement mediat  
 C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology  
 C:Keywords: glycoprotein; transmembrane protein  
 F:1-36/Domain: signal sequence #status predicted <SIG>  
 F:37-488/Product: complement regulatory protein, 512 antigen #status predicted <MAT>  
 F:38-94/Domain: complement factor H repeat homology <FH1>  
 F:98-156/Domain: complement factor H repeat homology <FH2>  
 F:161-227/Domain: complement factor H repeat homology <FH3>  
 F:233-288/Domain: complement factor H repeat homology <FH4>  
 F:294-352/Domain: complement factor H repeat homology <FH5>  
 F:357-413/Domain: complement factor H repeat homology <FH6>  
 F:421-444/Domain: transmembrane #status predicted <TM>  
 F:247,331,346,450,482,483/Binding site: carboxylate (Asn) (covalent) #status predict

Query Match 17.2%; Score 323; DB 2; Length 497;  
 Best Local Similarity 28.3%; Pred. No. 5.6e-16;  
 Matches 95; Conservative 55; Mismatches 144; Indels 42; Gaps 21;

QY 21 DQTYPEGTQAIYKCRPGYRSLGNVIMCRKGEVWALNPLKRCQKRPCGHPDTPPGTTL 80  
 Db 55 ESTPVGTSLSKYECPYIKRPSITCEVNSVWT--SPQDVCIKQCEPLDPQNGIVH 112  
 QY 81 TGGNVEYGVKAVYTCNBSGOLLGELINRECDTGD---GWNDIPICEVVK-LPVTAPEN 136  
 Db 113 --NTDIRFGSSITTYTCNBSGOLLGSSAMCTISDQSYAWAEAPICISICEIPSPISIP 169  
 QY 137 GKIYSSAMEPDR-YHFGQAVRFVCSNG-----YKIEGDEEMHC-----DDGFWSEKXP 185  
 Db 170 GDFFS-----PNRDEPHYGMVTVYQCNLDARGKLFNLVGEPSIHCTSIDQGVWMSGPPP 225  
 QY 186 KCEYIS-CKSPDYINGSPISQ-KIYKENERFOYKCNMGYEYSEKGDVCTE-SGWRP-L 241  
 Db 226 QCIETLNKCTPPHENVAVTYSKNSLSFLRDMVEFRQDGMKGDSSVYCRSLNRMWEPOL 285  
 QY 242 PSC-EKESCDN--PYIPNGD-YSPLRKHTGDEITYQCRNGFPYATRGNTAK-CTSTG- 295  
 Db 286 PSCFKKSCGAFGLPNGHVFVPOUN--OLGAKVTFVCTGTG--OLKGNSSSHCVLDGV 341  
 QY 296 --W-IPAPRCTLKPDYDIPDKIKHGLYHENMRRPYF 327  
 Db 342 ESTMNSVVPYCEQVICKLPQDMSGFQGLQMKKDY 377

## RESULT 15 I46001

C4b-binding protein alpha chain - bovine  
 C:Species: Bos primigenius taurus (cattle)  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 A:Accession: I46001; S43190  
 R:Hillarp, A.; Thern, A.; Dahlback, B.  
 J. Immunol. 153, 4190-4199, 1994  
 A:Title: Bovine C4b binding protein. Molecular cloning of the alpha- and beta-chains  
 A:Reference number: I46001; MUID:95015909  
 A:Accession: I46001

A; Status: preliminary; translated from GB/EMBL/DBJ

A; molecule type: mRNA

A;Residues: 1-610 &lt;HIL&gt;

A; Cross-references: EMBL:Z31693; NID:g469117; PIDN:CAA83498.1; PID:g469118

C3; Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology <FH1> C5; 107/8; complement factor H repeat homology <FH1>

F;50-107/Domain: complement factor H repeat homology <FH1>

F;112-169/Domain:	complement factor H repeat homology <FH2>
E;174-234/Domain:	complement factor H repeat homology <FH3>

F;174-234/Domain:	complement factor H repeat	<FH3>
E;238-264/Domain:	complement factor H repeat	<FH3>

F; 239-294/Domain:	complement factor H repeat	<EH4>
F; 239-294/Domain:	complement factor H repeat	<EH4>
F; 299-363/Domain:	complement factor H repeat	<EH5>

F: 365-425/Domain:	complement factor H repeat homology <FH6>
F: 299-362/Domain:	complement factor H repeat homology <FH5>

F:429-483/Domain: complement factor H repeat	homology <FHR>
F:366-425/Domain: complement factor H repeat	homology <FHS>

F:487-541/Domain: complement factor H repeat	<FH8>
F:429-483/Domain: complement factor H repeat	<FH7>
F:487-541/Domain: complement factor H repeat	<FH8>

EF48/-34L/Domain: complement factor D repeat homolog -1982

Query Match	17.18;	Score 320.5;	DB 1;	Length 610;
Query Match	27.79;	Score 16.15;	DB 1;	Length 610;

Best Local Similarity 27.7%; Pred. No. 1.1e-15;  
 Watcher 00: Conservative 54: Mismatches 142

Matches	99;	Conservative	54;	Mismatches	142;	Indels	63;	Gaps	22;
---------	-----	--------------	-----	------------	------	--------	-----	------	-----

```

0Y      6 LPPRNTELLIGMSWDQJYPPGTOAIYCRGGRY--SIGVNIIMCRGGEVWALNPLRKCQ 63
Db      52 IPPYLDFAEPPIINELNETREFTGTLRLYCRGRYIISAKNFELIODGTDNM---KYKEFCV 108
0Y      64 KRQCGHPGDPTRGFTLLGAVN-----FEYVKAIVYTCNEGYOLLGEIN--YRECDTG-- 115
Db      109 KRCENPGE-----LANGOYIAKTDYBSFSELEFCSSEBYVILIGSANEYQLODKGY 161
0Y      116 WNDPIPIEYVVKCLPVTAPENGRKIVSSAMEBDREYHGOAVAFVCSNGYIEGDEMHCS 175
Db      162 WSDPLQCIACNEPPIITSNGR--HNGGDBD-FYTVGSSTVYSCDRBFSMLGRKASTISR 218
0Y      176 DD----GFWSEKRPCKVEISCKSPDYINGSPISOKI-----IYKENBRPOYKCMQET 225
Db      219 VENKTIQVSPSPCKRKVICYQPVYKDG---KITSFEPITYOOSIYVAKNCKFRL 273
0Y      226 SERGAV--C-IESGWR-PLPSCSEKSC-DNRYIPNDYISPLRIKHRTGDE----- 271Y
Db      274 --EGDSLHICEDANSWNBPPTCELNGCLGLPHLPHALME--RKHDTQTEQOYVYDIGV 329
0Y      272 IYOCENGYPATRG--NPAKCTST--GWIPAPRCLTLPEDYDIDIKHG--IYHEMMRRP 325
Db      330 LSTKCHFGKAPETDGPITYTOSNLNMFPIYECKEVCCPEPNLNNYSITLH---RRP 384

```

Search completed: November 21, 2000, 16:47:40  
Job time: 369 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2000, 16:47:33 ; Search time 47.7 seconds  
(without alignments)  
352.557 Million cell updates/sec

Title: US-09-316-163-10

Perfect score: 1497

Sequence: 1 EDCNLEPPRRNTEILTGSWS.....EKSCDNPYIPNGDYSPLRIK 265

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters: 182106

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: p1r1: \*  
2: p1r2: \*  
3: p1r3: \*  
4: p1r4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	1497	100.0	449	1 NBHHS	complement factor
2	1497	100.0	1231	1 NBHHS	complement factor
3	1024	68.4	1234	1 NBHHS	complement factor
4	699	46.7	669	2 S6551	factor H - bovine
5	440	29.4	1053	2 S4619	probable complemen
6	332.5	22.2	360	2 T42921	complement control
7	303	20.2	597	1 NBHUC4	C4b-binding protel
8	302	20.2	597	1 S53711	C4BP alpha chain p
9	301.5	20.1	263	1 WMYZSP	apolipoprotein h
10	301	20.1	676	2 A45900	complement C3b rec
11	298.5	19.9	2489	2 I73012	complement C3b/C4b
12	290.5	19.4	2014	2 I36936	complement recepto
13	289.5	19.3	497	2 JC2054	complement regulat
14	284.5	19.0	263	1 C36838	hypothetical prote
15	284.5	19.0	263	2 T28450	hypothetical prote
16	284.5	19.0	482	2 A34924	complement C3b/C4b
17	283.5	18.9	263	2 B72152	B18L protein - var
18	277.5	18.5	560	2 T16833	hypothetical prote
19	276.5	18.5	558	2 S57953	C4BP protein alpha
20	275	18.4	469	1 NBHSC4	C4b-binding protel
21	274.5	18.3	349	2 G02913	sperm CD46 - human
22	274.5	18.3	369	2 I57998	membrane cofactor
23	274	18.3	377	2 I54479	membrane cofactor
24	274	18.3	384	2 S01896	membrane cofactor
25	273.5	18.3	433	2 A30550	complement C3b/C4b
26	272	18.2	610	2 A35046	E-selectin precurs
27	270.5	18.1	362	2 JC5194	membrane cofactor
28	270.5	18.1	369	2 JC5138	membrane cofactor
29	270.5	18.1	440	2 A43519	complement recepto

30	266	17.8	482	2 JC5092	E-selectin - pig
31	265.5	17.7	610	1 T46001	C4b-binding protei
32	264.5	17.7	302	1 WMBE1E	secretory compleme
33	264.5	17.7	360	1 WMBE2E	membrane-bound com
34	262	17.5	381	1 B26359	decay-accelerating
35	262	17.5	440	2 A26359	decay-accelerating
36	260.5	17.4	768	2 A42755	P-selectin precurs
37	260	17.4	612	2 S23174	endothelial leukoc
38	257.5	17.2	1025	1 A43526	complement C3d/Eps
39	257	17.2	579	2 A56740	sperm-egg recognit
40	256.5	17.1	618	2 B42755	E-selectin recognit
41	255	17.0	452	2 A35068	complement factor
42	254	17.0	768	2 I53821	P-selectin - rat
43	253	16.9	345	1 NBMS	apolipoprotein h p
44	252.5	16.9	1091	1 PL0009	complement C3d/Eps
45	252	16.8	340	2 I56234	decay-accelerating

## ALIGNMENTS

RESULT 1  
NBHHS  
complement factor H precursor, short splice form - human  
N: Alternate names: complement factor H-related protein; complement protein H  
C: Species: Homo sapiens (man)  
C: Date: 31-Dec-1993 #sequence-revision 23-Feb-1996 #text-change 22-Jun-1999  
C: Accession: S03013; B60238; A27877; A61103; A26505; S10479  
R: Ripoché, J.; Day, A.J.; Harris, T.J.R.; Sim, R.B.  
Biochem. J. 249, 593-602, 1988  
A: Title: The complete amino acid sequence of human complement factor H.  
A: Reference number: S00254; MUID: 88134059  
A: Accession: S03013  
A: Molecule type: mRNA  
A: Residues: 1-449 <RIP>  
A: Cross-references: EMBL:X07523; EMBL:Y00716; NID:932492; PIDN:CAA30403.1; PID:975807  
A: Note: part of this sequence, including the amino end of the mature protein was conf  
A: Note: 402-Tyr was also found  
R: Estalier, C.; Schwaible, W.; Dierich, M.; Weiss, E.H.  
Eur. J. Immunol. 21, 799-802, 1991  
A: Title: Human complement factor H: two factor H proteins are derived from alternativ  
A: Reference number: A60238; MUID: 91184292  
A: Accession: B60238  
A: Status: not compared with conceptual translation  
A: Molecule type: mRNA  
A: Residues: 1-33;434-449 <EST>  
A: Note: Only portions of this 1.8 kilobase mRNA were sequenced  
R: Schultz, T.F.; Schwaible, W.; Stanley, K.K.; Weiss, E.; Dierich, M.P.  
Eur. J. Immunol. 16, 1351-1355, 1986  
A: Title: Human complement factor H: isolation of cDNA clones and partial cDNA sequence  
A: Reference number: A27877; MUID: 87054207  
A: Accession: A27877  
A: Molecule type: mRNA  
A: Residues: '11',55-401,'Y',403-449 <SCH>  
A: Note: an additional nucleotide present within the codon for Glu-310 was thought to  
R: Schwaible, W.; Zwirner, J.; Schulz, T.F.; Linke, R.P.; Dierich, M.P.; Weiss, E.H.  
Eur. J. Immunol. 17, 1485-1489, 1987  
A: Title: Human complement factor H: expression of an additional truncated gene produc  
A: Reference number: A61103; MUID: 88055295  
A: Accession: A61103  
A: Status: not compared with conceptual translation  
A: Molecule type: mRNA  
A: Residues: 17-76 <SC2>  
A: Note: This is a partial sequence of an alternatively spliced 1.8 kilobase mRNA that  
R: Sim, R.B.; Discipio, R.G.  
Biochem. J. 205, 285-293, 1982  
A: Title: Purification and structural studies on the complement-system control protein  
A: Reference number: A26505; MUID: 83048213  
A: Accession: A26505  
A: Molecule type: protein  
A: Residues: 19-20,'Q',22-29,'V',31-33,'Q',35 <SIM>  
R: Barlow, P.N.; Norman, D.G.; Steinkasserer, A.; Horne, T.J.; Pearce, J.; Driscoll, P  
Biochemistry 31, 3626-3634, 1992



	C:Comment:	Alternative transcripts of 4.3, 1.8, and 1.4 kilobases are expressed in liver
	C:Genetics:	<HF1>
	A:Gene:	GDB:HF1; HF
	A:Cross-references:	GDB:120041; OMIM:134370
	A:Map position:	Iq32-Iq32
	C:Genetics:	<HF2>
	A:Gene:	GDB:HF2; HF
	A:Cross-references:	GDB:129095
	A:Map position:	Iq32-Iq32
	A>Note:	The correspondence between the two loci and the sequences indicated is unclear;
	C:function:	
	A:Description:	a cofactor in the inactivation of C3b by serine proteinase I; also increas-
	A:Pathway:	complement alternate pathway
	C:Superfamily:	complement factor H; complement factor H repeat homology
	C:Keywords:	alternative splicing; complement status predicted <SIG>
	F:11-18/Domains:	signal sequence #status experimental <MP7>
	F:19-1228/Product:	complement factor H, short splice form #status experimental <MAT>
	F:21-80/Domains:	complement factor H repeat homology <FH01>
	F:85-141/Domains:	complement factor H repeat homology <FH02>
	F:146-205/Domains:	complement factor H repeat homology <FH03>
	F:210-262/Domains:	complement factor H repeat homology <FH04>
	F:246-248/Region:	cell attachment (R-G-D) motif
	F:326-330/Domains:	complement factor H repeat homology <FH05>
	F:329-385/Domains:	complement factor H repeat homology <FH06>
	F:389-442/Domains:	complement factor H repeat homology <FH07>
	F:448-505/Domains:	complement factor H repeat homology <FH08>
	F:509-564/Domains:	complement factor H repeat homology <FH09>
	F:568-623/Domains:	complement factor H repeat homology <FH10>
	F:630-684/Domains:	complement factor H repeat homology <FH11>
	F:691-744/Domains:	complement factor H repeat homology <FH12>
	F:753-803/Domains:	complement factor H repeat homology <FH13>
	F:811-864/Domains:	complement factor H repeat homology <FH14>
	F:870-926/Domains:	complement factor H repeat homology <FH15>
	F:931-984/Domains:	complement factor H repeat homology <FH16>
	F:989-1043/Domains:	complement factor H repeat homology <FH17>
	F:1048-1102/Domains:	complement factor H repeat homology <FH18>
	F:1109-1163/Domains:	complement factor H repeat homology <FH19>
	F:1167-1228/Domains:	complement factor H repeat homology <FH20>
	F:121-66, 52-80, 85-129, 114-141, 146-192, 178-205, 210-251, 237-262, 267-309, 294-320, 325-374, 357-403, 408-439, 444-475, 480-511, 516-547, 552-583, 588-619, 624-655, 660-691, 696-727, 732-763, 768-800, 805-836, 841-872, 877-908, 913-944, 949-980, 985-1016, 1021-1052, 1057-1088, 1093-1124, 1129-1160, 1165-1196, 1201-1232, 1237-1268, 1273-1304, 1309-1340, 1345-1376, 1381-1412, 1417-1448, 1453-1484, 1489-1520, 1525-1556, 1561-1592, 1597-1628, 1633-1664, 1669-1700, 1705-1736, 1741-1772, 1777-1808, 1813-1844, 1849-1880, 1885-1916, 1921-1952, 1957-1988, 1993-2024, 2029-2060, 2065-2096, 2101-2132, 2137-2168, 2173-2204, 2209-2240, 2245-2276, 2281-2312, 2317-2348, 2353-2384, 2389-2420, 2425-2456, 2461-2492, 2497-2528, 2533-2564, 2569-2600, 2605-2636, 2641-2672, 2677-2708, 2713-2744, 2749-2780, 2785-2816, 2821-2852, 2857-2888, 2893-2924, 2929-2960, 2965-2996, 3001-3032, 3037-3068, 3073-3104, 3109-3140, 3145-3176, 3181-3212, 3217-3248, 3253-3284, 3289-3320, 3325-3356, 3361-3392, 3397-3428, 3433-3464, 3469-3500, 3505-3536, 3541-3572, 3577-3608, 3613-3644, 3649-3680, 3685-3716, 3721-3752, 3757-3788, 3793-3824, 3829-3860, 3865-3896, 3901-3932, 3937-3968, 3973-4004, 4009-4040, 4045-4076, 4081-4112, 4117-4148, 4153-4184, 4189-4220, 4225-4256, 4261-4292, 4297-4328, 4333-4364, 4369-4400, 4405-4436, 4441-4472, 4477-4508, 4513-4544, 4549-4580, 4585-4616, 4621-4652, 4657-4688, 4693-4724, 4729-4760, 4765-4796, 4801-4832, 4837-4868, 4873-4904, 4909-4940, 4945-4976, 4981-5012, 5017-5048, 5053-5084, 5089-5120, 5125-5156, 5161-5192, 5197-5228, 5233-5264, 5269-5300, 5305-5336, 5341-5372, 5377-5408, 5413-5444, 5449-5480, 5485-5516, 5521-5552, 5557-5588, 5593-5624, 5629-5660, 5665-5696, 5701-5732, 5737-5768, 5773-5804, 5809-5840, 5845-5876, 5881-5912, 5917-5948, 5953-5984, 5989-6020, 6025-6056, 6061-6092, 6097-6128, 6133-6164, 6169-6200, 6205-6236, 6241-6272, 6277-6308, 6313-6344, 6349-6380, 6385-6416, 6421-6452, 6457-6488, 6493-6524, 6529-6560, 6565-6596, 6601-6632, 6637-6668, 6673-6704, 6709-6740, 6745-6776, 6781-6812, 6817-6848, 6853-6884, 6889-6920, 6925-6956, 6961-6992, 6997-7028, 7033-7064, 7069-7100, 7105-7136, 7141-7172, 7177-7208, 7213-7244, 7249-7280, 7285-7316, 7321-7352, 7357-7388, 7393-7424, 7429-7460, 7465-7496, 7501-7532, 7537-7568, 7573-7604, 7609-7640, 7645-7676, 7681-7712, 7717-7748, 7753-7784, 7789-7820, 7825-7856, 7861-7892, 7897-7928, 7933-7964, 7969-8000, 8005-8036, 8041-8072, 8077-8108, 8113-8144, 8149-8180, 8185-8216, 8221-8252, 8257-8288, 8293-8324, 8329-8360, 8365-8396, 8401-8432, 8437-8468, 8473-8504, 8509-8540, 8545-8576, 8581-8612, 8617-8648, 8653-8684, 8689-872	

NEMSH  
 complement factor H precursor - mouse  
 M:Alternate names: protein beta-1-H  
 C:Species: Mus musculus (house mouse)  
 C:Date: 30-Sep-1987 (sequence\_revision 30-Sep-1987 #text\_change 22-Jun-1999  
 C:Accession: A26154; 149711; 149728  
 R:Kristensen, T.; Tack, B.F.  
 Proc. Natl. Acad. Sci. U.S.A. 83, 3963-3967, 1986  
 A:Title: Murine protein H is comprised of 20 repeating units, 61 amino acids in length  
 A:Reference number: A26154; MUID:86233353  
 A:Accession: A26154  
 A:Molecule type: mRNA  
 A:Residues: 1-1234 <RRI>  
 A:Cross-references: GB:M12660; NID:g193724; PIDN:AAA37759.1; PID:g387181  
 R:Natsunome-Sakii, S.; Nonaka, M.; Nonaka, M.; Harada, Y.  
 J. Immunol. 144, 358-362, 1990  
 A:Title: Demonstration of an unusual allelic variation of mouse factor H by the comp  
 A:Reference number: 149711; MUID:90111033  
 A:Accession: 149711  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-18 <RES>  
 A:Cross-references: GB:M31979; NID:g193726; PIDN:AAA37762.1; PID:g193729  
 R:Munoz-Canoves, P.; Tack, B.F.; Vlk, D.P.  
 Biochemistry 28, 9891-9897, 1989  
 A:Title: Analysis of complement factor H mRNA expression: Dexamethasone and IFN-gamma  
 A:Reference number: 149728; MUID:90148935  
 A:Accession: 149728  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-19 <RE2>  
 A:Cross-references: GB:J02891; NID:g193805; PIDN:AAA37795.1; PID:g553926  
 C:Comment: Two codominant alleles of factor H are present in mice.  
 C:Comment: Factor H functions as a cofactor in the inactivation of C3b by serine prote  
 (C5 convertase) in the alternative complement pathway.  
 C:Genetics:  
 A:Map position: 1  
 C:Superfamily: complement factor H; complement factor H repeat homology  
 C:Keywords: complement alternate pathway; duplication; glycoprotein; plasma  
 F:1-18/Domain: signal sequence #status predicted <SIG>  
 F:19-1234/Product: complement factor H #status predicted <MP7>  
 F:21-80/Domain: complement factor H repeat homology <FH01>  
 F:85-141/Domain: complement factor H repeat homology <FH02>  
 F:146-205/Domain: complement factor H repeat homology <FH03>  
 F:210-262/Domain: complement factor H repeat homology <FH04>  
 F:246-248/Region: cell attachment (R-G-D) motif  
 F:267-320/Domain: complement factor H repeat homology <FH05>  
 F:325-385/Domain: complement factor H repeat homology <FH06>  
 F:387-442/Domain: complement factor H repeat homology <FH07>  
 F:448-505/Domain: complement factor H repeat homology <FH08>  
 F:509-564/Domain: complement factor H repeat homology <FH09>  
 F:569-622/Domain: complement factor H repeat homology <FH10>  
 F:629-683/Domain: complement factor H repeat homology <FH11>  
 F:690-743/Domain: complement factor H repeat homology <FH12>  
 F:752-802/Domain: complement factor H repeat homology <FH13>  
 F:808-861/Domain: complement factor H repeat homology <FH14>  
 F:867-931/Domain: complement factor H repeat homology <FH15>  
 F:933-989/Domain: complement factor H repeat homology <FH16>  
 F:994-1048/Domain: complement factor H repeat homology <FH17>  
 F:1053-1107/Domain: complement factor H repeat homology <FH18>  
 F:1114-1168/Domain: complement factor H repeat homology <FH19>  
 F:1172-1233/Domain: complement factor H repeat homology <FH20>  
 F:121-66,550,85-129,114-141,146-192,178-205,210-251,237-262,267-309,294-320,325-374,4  
 F:802,808-850,836-861,867-920,906-931,936-978,964-989,994-1037,1023-1048,1053-1096,10  
 F:676,721,773,801,1030,1061,1225/Binding site: carbohydrate (Asn) (covalent) #status

Db 19 EDCKPPRENSSELLSGWSBQLPEGTQATYKCRPGVRLGTIVKCKNGKWAANSR 78  
 QY 61 KQKRCPCGHGPDTPFGFTLTGCVNFEYGVKAVYTCNBGVLCEINRECDTDGWTNDI 120  
 Db 79 ICRKKPCGHGPDTPFGSGFRLAVGSGQFEFGAKVYTCDDGYOLLGEIDYREGGAGWINDI 138  
 QY 121 PICGVNVCPLVTPAENKITYSSAMEPRVHFGAVFVCSGKRTGDEMHGSDGQFW 180  
 Db 139 PLCEVNVCLPVTLENERIVSGAETDQYEFVGVVRECSGFKIRGKHTEHCSENGW 198  
 QY 181 SKERKCVETISCKSPDVINSPIQKITYKENERFOYKCNMGYSEYSGDAVCTESGWRP 240  
 Db 199 SNEKPRCVETLCTPRVENGGINVKPYKENERHYKCKAGVYPKERGAVCTGSGMS 258  
 QY 241 LPSCEEKSCDNPYTPNDYSPLRT 264  
 Db 259 QPCEERKCSPPYLLNGIYTPHRI 282

## RESULT 4

Factor H - bovine (fragment)  
 C:Species: Bos primigenius taurus (cattle)  
 C>Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
 C:Accession: S65551  
 R:Soames, C.J.; Day, A.J.; Sim, R.B.  
 Biochem. J. 315, 523-531, 1996  
 A>Title: Prediction from sequence comparisons of residues of factor H involved in the  
 A:Reference number: S65551; MUID:96202005  
 A:Accession: S65551  
 A>Status: preliminary: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-669 <SOA>  
 A:Cross-references: GB:X98697; NID:g1419423; PIDN:CAAG7257.1; PID:g1419424  
 C:Superfamily: complement factor H; complement factor H repeat homology  
 F:55-114/Domain: complement factor H repeat homology <FHR1>  
 F:296-149/Domain: complement factor H repeat homology <FHR2>  
 F:355-412/Domain: complement factor H repeat homology <FHR3>  
 F:416-471/Domain: complement factor H repeat homology <FHR4>  
 F:476-530/Domain: complement factor H repeat homology <FHR>  
 F:538-592/Domain: complement factor H repeat homology <FHR5>  
 F:599-651/Domain: complement factor H repeat homology <FHR6>

Query Match 46.7%; Score 699; DB 2; Length 669;  
 Best Local Similarity 63.4%; Pred. No. 1.2e-45;  
 Matches 118; Conservative 26; Mismatches 42; Indels 0; Gaps 0;

QY 76 GTFLLTGNNVEYGVKAVYTCNEGQVLLGEINRECDTDGWTNDIPICEVVKCLPTAPE 135  
 Db 3 GSPLIAGNQEYFAYKAVYTCDEGYQWGMNFRCDTNGTNDIPICEVVKCLPTAPE 62  
 QY 136 NGKIVSSAMEPRDEYHFGQAVRVCNSGYKTEGDEMHGSDGFWSKERKPCVETISCKSP 195  
 Db 63 NGKIFSPALIEDPDYTYGQVAVFECNSGYMLDGPQIHCSAGVWSAATPCVETIFCKXP 122  
 QY 196 DVINGSPISQKITYKENERFOYKCNMGYSEYSGDAVCTESGWRPPLPSCEKSCDNPYIP 255  
 Db 123 VILNGQAVLPKATYKQNERVOYRCAAGFEIGQRDYVCTKSGMPPAPICITITCDPRIP 182  
 QY 256 NGDYSP 261  
 Db 183 NGVYRP 188

## RESULT 5

Factor H - bovine (fragment)  
 C:Species: Parabrax neblifer (barred sand bass)  
 C>Date: 19-Mar-1997 #sequence\_revision 06-Jun-1997 #text\_change 08-Oct-1999  
 C:Accession: S46199; S77894  
 R: Dahmen, A.; Kaidoh, T.; Zupfel, P.F.; Gligli, I.  
 Biochem. J. 301, 391-397, 1994

A>Title: Cloning and characterization of a cDNA representing a putative complement-re  
 A:Reference number: S46199; MUID:94318039  
 A:Accession: S46199  
 A:Molecule type: mRNA  
 A:Residues: 1-1053 <DAH1>  
 A:Cross-references: EMBL:L21703; NID:g639894; PIDN:AAA92556.1; PID:g639895  
 A:Experimental source: liver  
 A:Accession: S77894  
 A:Molecule type: protein  
 A:Residues: 526-532, 'X', 534-537; 809-817, 'X', 819-826 <DAH2>  
 C:Genetics:  
 A:Gene: snl  
 C:Superfamily: complement factor H repeat homology  
 C:Keywords: glycoprotein  
 F:89-145/Domain: complement factor H repeat homology <FH01>  
 F:334-389/Domain: complement factor H repeat homology <FH02>  
 F:450-502/Domain: complement factor H repeat homology <FH03>  
 F:569-624/Domain: complement factor H repeat homology <FH03>  
 F:682-738/Domain: complement factor H repeat homology <FH04>  
 F:743-802/Domain: complement factor H repeat homology <FH04>  
 F:935-989/Domain: complement factor H repeat homology <FH05>  
 F:993-1052/Domain: complement factor H repeat homology <FH06>

Query Match 29.4%; Score 440; DB 2; Length 1053;  
 Best Local Similarity 36.2%; Pred. No. 8.6e-26;  
 Matches 89; Conservative 35; Mismatches 108; Indels 14; Gaps 6;

QY 21 DQYPEGTQATYKCRPGVRLGTIVKCKNGKWAANSR 78  
 Db 47 EASYPGHQVAVGCVNGVS--GPFKLVCEGKMETRG--AKCOPRSGHPGDAQFADPHL 102  
 QY 81 TGVNFEYGVKAVYTCNBGVLCEINRECDTDGWTNDIPICEVVKCLPTAPENGKIY 140  
 Db 103 AEGNDFEFGSKVYVTCQGYQWVSINIRKCVAGGWDGVVVCSSQOPLIHVNNQVI 162  
 QY 141 SSAMEPRDEYHFGQAVRVCNSGYKTEGDEMHGSDGFWSKERKPCVETISCKSPDVIN 199  
 Db 163 GG----PEARFNGVNRSCSRSEILIDGSPELICDEGMDSGVPRKCAITCAIPLEN 218  
 QY 200 GSPISQKITYKENERFOYKCNMGYSEYSGDAVCTESG----WRPLPSCEKSCDNPYIP 255  
 Db 219 GNVGAIREYKENDVLAHEDCRAFKHIDR-PSTCIKQIKAKWSPTPLCESIKRLTIMD 277  
 QY 256 NGDYSP 261  
 Db 278 GTRYEP 283

RESULT 6  
 T42921  
 complement control protein homolog ccph - ateline herpesvirus 3 (strain 73)

C:Species: ateline herpesvirus 3  
 A:Variety: strain 73  
 C>Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 04-Mar-2000  
 C:Accession: T42921  
 R:Albrecht, J.C.; Fleckenstein, B.  
 Submitted to the EMBL Data Library, August 1998  
 A:Description: Primary structure of the herpesvirus ateline genome.  
 A:Reference number: Z22274  
 A:Accession: T42921  
 A:Molecule type: DNA  
 A:Status: preliminary: translated from GB/EMBL/DBJ  
 A:Cross-references: EMBL:AF083424; PIDN:AA095530.1  
 A:Experimental source: strain 73  
 C:Superfamily: herpesvirus complement control protein; complement factor H repeat hom

Query Match 22.2%; Score 332.5; DB 2; Length 360;  
 Best Local Similarity 34.9%; Pred. No. 3.8e-18;  
 Matches 96; Conservative 32; Mismatches 106; Indels 41; Gaps 18;

[illegible]

A:Accession: A93134  
A:Molecule type: protein  
A:Residues: 49-81 <CHI>  
A:Note: this paper reports amino-terminal sequences of the intact protein and of a nu  
R:Hessing, M.; Kanfers, D.; Takeya, H.; van't Veer, C.; Hackeng, T.M.; Iwanaga, S.; B  
PDB Lett. 317, 228-232, 1993  
A:Title: The region Ser(333)-Arg(336) of the alpha-chain of human C4b-binding protein  
A:Reference number: S29492; MUID:93146164  
A:Accession: S29492  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 381-404 <HES>  
R:Suzuki, K.; Nishio, J.  
J. Biol. Chem. 263, 17034-17039, 1988  
A:Title: Binding site for vitamin K-dependent protein S on complement C4b-binding prot  
A:Reference number: A31785; MUID:89034204  
A:Accession: A31785  
A:Molecule type: protein  
A:Residues: 495-505, 'Y', 507-510, 'X', 512-515 <SUZ>  
A:Note: this peptide appears to bind protein S  
R:Danilack, B.; Smith, C.A.; Muller-Eberhard, H.J.  
Proc. Natl. Acad. Sci. U.S.A. 80, 3461-3465, 1983  
A:Title: Visualization of human C4b-binding protein and its complexes with vitamin K-  
A:Reference number: A59350; MUID:85221615  
A:Contents annotation: electron microscopy; three-dimensional structure; ligand bind  
R:Aso, T.; Okamura, S.; Matsuguchi, T.; Sakamoto, N.; Sata, T.; Niho, Y.  
Biochem. Biophys. Res. Commun. 174, 222-227, 1991  
A:Title: Genomic organization of the alpha chain of the human C4b-binding protein gene  
A:Reference number: I52244; MUID:91113199  
A:Accession: I52244  
A:Status: translation not shown; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-597 <ASO>  
A:Cross-references: GB:M62486; NID:9190498; PIND:AAA36506.1; PID:9190500  
C:Comment: C4BP controls the classical pathway of complement activation. It binds as  
the degradation of the C4bC2a complex (C3 convertase) by dissociating the complement  
C:Comment: C4BP occurs in plasma in two forms, both of which bind complement fragment  
I and II bonds. The beta chain binds the vitamin K-dependent plasma protein S. A minor  
actors V and VIII.  
C:Comment: The molecule has a central body supporting seven tentacles (alpha chains),  
C:Genetics:  
A:Gene: GDB:C4BPA  
A:Cross-references: GDB:120568; OMIM:120830  
A:Map position: 1q32-1q32  
A:Introns: 48/1; 110/1; 143/2; 172/1; 236/1; 297/1; 362/1; 425/1; 482/1; 540/3  
C:Complex: octomer of seven alpha chains and one beta chain  
C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology  
C:Keywords: acute phase; chylomicron; complement pathway; duplication; glycoprotein;  
F:1-48/Domain: signal sequence #status predicted <SIG>  
F:49-597/Product: C4b-binding protein alpha chain #status predicted <MNT>  
F:50-108/Domain: complement factor H repeat homology <CHI>  
F:113-170/Domain: complement factor H repeat homology <FHI>  
F:175-234/Domain: complement factor H repeat homology <FHI>  
F:239-294/Domain: complement factor H repeat homology <FHI>  
F:299-360/Domain: complement factor H repeat homology <FHI>  
F:364-422/Domain: complement factor H repeat homology <FHI>  
F:381-404/Region: complement C4b binding #status predicted  
F:426-480/Domain: complement factor H repeat homology <FHI>  
F:484-538/Domain: complement factor H repeat homology <FHI>  
F:521-506,528/Binding site: carbohydrate (Asn) (covalent) #status experimental



Query Match	20.1%;	Score 301;	DB 2;	Length 676;
Best Local Similarity	26.0%;	Pred. No. 1.8e-15;		
Matches 82; Conservative	51;	Mismatches 94;	Indels 88;	Gaps 16

## RESULT 11

Residues: 1-683, 'X', 685-894, 'A', 896-1000, 1451-1471, 'X', 1473-2489 <VIK2>

A:Cross-references: GB:LI7418; NID:g306678; PID:g306680  
R:Wong, W.W.; Cahill, J.M.; Rosen, M.D.; Kennedy, C.A.; Bonaccio, E.T.; Morris, M.J.;  
J. Exp. Med. 169, 847-863, 1989  
A:Title: Structure of the human C3b gene. Molecular basis of the structural and quant  
A:Reference number: A47602; MUID:89176869  
A:Accession: A47602  
A:Molecule type: DNA  
A:Residues: 1-41 <MON>  
A:Cross-references: GB:X14893  
R:Hourcade, D.; Miesner, D.R.; Atkinson, J.P.; Holers, V.M.  
J. Exp. Med. 168, 1255-1270, 1988  
A:Title: Identification of an alternative polyadenylation site in the human C3b/C4b r  
type 1.  
A:Reference number: S03291; MUID:89010527  
A:Accession: S03291  
A:Molecule type: mRNA  
A:Residues: 26-584 <HOU>  
A:Cross-references: EMBL:X14362; NID:g30197; PID:g736240  
A:Experimental source: clone CRI-4  
R:Klickstein, L.B.; Barrow, T.J.; Miletic, V.; Rabson, L.D.; Smith, J.A.; Fearon, D.T.  
J. Exp. Med. 168, 1699-1717, 1988  
A:Title: Identification of distinct C3b and C4b recognition sites in the human C3b/C4  
A:Reference number: S03843; MUID:89035992  
A:Accession: S03843  
A:Status: translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-894, 'A', 896-1000, 1451-2064, 'T', 2066-2276, 'P', 2278-2299, 'H', 2301-2325, 'T'  
A:Cross-references: EMBL:Y00816; NID:g30185; PID:g30186  
R:Klickstein, L.B.; Wong, W.W.; Smith, J.A.; Wels, J.H.; Wilson, J.G.; Fearon, D.T.  
J. Exp. Med. 165, 1095-1112, 1987  
A:Title: Human C3b/C4b receptor (CRL). Demonstration of long homologous repeating dom  
A:Reference number: A28507; MUID:87168191  
A:Accession: A28507  
A:Molecule type: mRNA  
A:Residues: 953-1221, 'F', 1224-2064, 'T', 2066-2276, 'P', 2278-2299, 'H', 2301-2325, 'T', 2332  
A:Cross-references: GB:X05309; NID:g30196; PID:CA28933.1; PID:g809019  
R:Mong, W.W.; Klickstein, L.B.; Smith, J.A.; Wels, J.H.; Fearon, D.T.  
Proc. Natl. Acad. Sci. U.S.A. 82, 7711-7715, 1985  
A:Title: Identification of a partial cDNA clone for the human receptor for complement  
A:Reference number: A94073; MUID:86067975  
A:Accession: A24748  
A:Molecule type: mRNA  
A:Residues: 311-333, 729-745, 831-845 <MO2>  
A:Cross-references: GB:M1569; NID:g180991; PID:AAA52297.1; PID:g180995; GB:M1517;  
C:Genetics:  
A:Gene: GDB:CR1; CD35  
A:Cross-references: GDB:119800; OMIM:120620  
A:Map position: 1q32-1q32  
A:Introns: 41/1; 101/1; 134/2; 163/1; 296/1; 389/2; 418/1; 491/1; 551/1; 584/2  
; 1484/2; 1513/1; 1646/1; 1705/1; 1739/2; 1768/1; 1844/1; 1904/1; 1937/2; 1966/1; 209  
C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology  
C:Keywords: duplication; glycoprotein; transmembrane protein  
F:1-41/Domain: signal sequence #status predicted <SIG>  
F:42-2489/Product: complement C3b/C4b receptor, membrane-bound form #status predicted  
F:42-558/Product: complement C3b/C4b receptor, secreted form #status predicted <MATS>  
F:104-99/Domain: complement factor H repeat homology <FH01>  
F:166-232/Domain: complement factor H repeat homology <FH02>  
F:238-293/Domain: complement factor H repeat homology <FH03>  
F:297-353/Domain: complement factor H repeat homology <FH04>  
F:358-416/Domain: complement factor H repeat homology <FH05>  
F:421-487/Domain: complement factor H repeat homology <FH06>  
F:493-549/Domain: complement factor H repeat homology <FH07>  
F:554-611/Domain: complement factor H repeat homology <FH08>  
F:616-682/Domain: complement factor H repeat homology <FH09>  
F:688-743/Domain: complement factor H repeat homology <FH10>  
F:747-803/Domain: complement factor H repeat homology <FH11>  
F:808-866/Domain: complement factor H repeat homology <FH12>  
F:871-937/Domain: complement factor H repeat homology <FH13>  
F:943-999/Domain: complement factor H repeat homology <FH14>  
F:1004-1061/Domain: complement factor H repeat homology <FH15>  
F:1066-1132/Domain: complement factor H repeat homology <FH16>  
; 1138-1193/Domain: complement factor H repeat homology <FH17>





Tue Nov 21 16:57:28 2000

us-09-316-163-10.rpr

---

Page 10